

What kind of greenhouse treaty?

Events elsewhere should not blind us to the conference begun in Washington this week at which the first steps will be taken towards a treaty to avoid the effects of global change. There is a long way to go.

THERE could hardly have been a worse occasion than the present for the first meeting, in Washington this week, of the negotiators who are meant to produce a global warming treaty in the next sixteen months or so. Many of the governments whose support is essential have other things on their minds. The United States is preoccupied with the Gulf War, the Soviet Union with its own survival as an integral power and many of the other industrialized nations of the world with the signs of financial instability that abound — Britain and the United States are sliding into recession, Germany and Japan are fighting inflation by increasing interest rates. And who can expect Iraqi, Israeli and Saudi Arabian delegates to sit through a conference on global warming in a contemplative frame of mind? So nobody should be disappointed if little seems to happen in the next two weeks. It will be time enough for that if the next year goes by with nothing to report.

The immediate goal is not a treaty, but a framework within which a treaty can be negotiated. The timetable requires that to be done by the time of the conference planned for June next year. Given that it has taken nearly twenty years for the United States and the Soviet Union to fail to reach a bilateral agreement on strategic arms — that is the treaty that will not now be signed in Moscow this month — that has always been a tall order.

Those who ask that the treaty should include specific regulations for the control of greenhouse-gas emissions make it even less likely that a workable treaty could be ready in time. But the practical instrument, a treaty requiring its signatories to agree to act when the need for action has also been agreed, is also the wisest course to follow. The principle of the connection between the accumulation of greenhouse gases and the prospect of climatic change may be widely accepted, but the detailed consequences of climatic change, and the magnitude of their social consequences, can only be guessed at. To suggest otherwise, as greenhouse enthusiasts do, is irresponsible.

That is why close attention should be paid to proposals such as that of David Victor, on page 451 of this issue. Victor's appealing argument has several virtues, of which the chief is that it recognizes that the negotiation of a greenhouse treaty must be a continuing process, not a once and for all business.

Enthusiasm for writing into an instant treaty on global warming restrictions on emissions comparable with those

voluntarily undertaken by the governments of the European Communities may be well-meant, but would at this stage be unhelpful. For the European targets are insufficient to avoid increases of temperature such as those predicted by the numerical models. If the predictions are correct, there will be some chagrin among European voters persuaded that they have done their bit to avert global warming when they are told that even more is expected of them.

That is but one reason why it would be preferable that the international commitment to the abatement of greenhouse gases should be renewed at regular intervals, and in circumstances in which those concerned are made to strike equitable bargains with each other.

Issues of equity are bound to dominate the years ahead, but it would be folly to suppose that these could all be settled in advance. It is not merely that some means will have to be found for compensating developing countries for their avoidance of practices that have made rich countries prosperous, the burning of fossil fuel in particular, but that the rich countries will have legitimate grounds for quarrelling among themselves over issues such as the retail price of gasoline and the credit they are allowed for their dependence on nuclear electric generation. It is difficult enough to see how questions such as these will be resolved if and when the signs of climatic change become more definite than they are at present. Attempting to settle them in advance is a recipe for failure — and for costly error.

That is why the most important part of any framework negotiated in the next few months must be the arrangements made for alerting the signatories of a greenhouse treaty to the steps that next call for action. Victor's notion of a multidisciplinary secretariat based on that which has successfully administered the General Agreement on Tariffs and Trade (GATT) may not be the most compelling in the wake of the collapse of last December's trade negotiations, but that failure lies at politicians' doors. And there are now signs that the talks will resume.

What the greenhouse treaty needs is a body of competence at the analysis of the technical information that accumulates as well as the definition of goals that might realistically be achieved in the hard political world. When it takes much longer than sixteen months to found modest research institutes, the goals the treaty-writers have set themselves are quite breathtaking. □



Britain's science saga

British politicians are beginning to take an interest in the state of science. Not before time.

THIS week's debate in the House of Commons on the condition of the research councils whose chief function is the support of academic research is a landmark of a kind. It is a long time since such a question was raised in such a setting. And while there is no reason to expect that the new government will admit the errors of its predecessor's ways, there is at least a chance that it may be forced in future to be less cavalier than has been the custom.

But this week's debate (which will already have taken place by the time this issue is widely read) is narrow in scope, and chiefly concerned with the most recent cutting of the cake that is annually divided between the five research councils. It will be right and proper if the government has been given a drubbing over its decision to break with a precedent set by the then Sir Keith (now Lord) Joseph, when Secretary of State for Education and Science in the early 1980s, to publish the independent advice on which the cake-cutting was based. Joseph, hardly renowned for his attachment to open government, had soundly concluded that no harm would come of making the advice public even when he chose to override it. Of what is the new government afraid?

But the chances are that this week's debate will also have revealed the dangers inherent in arguments such as this over the budgetary problems. In particular, they create the impression that the problems of British science are money problems. Money matters, and more money would be beneficial in the present circumstances in which good projects are being held up or cancelled through budgetary accidents, but the serious troubles with British science are more deep-seated. The serious reason for alarm is that the research community is beginning to reap the painful reward for a decade's preoccupation by others with the supposed short-term benefits of research.

Not that British science is all dust and ashes. As the world knows, many laboratories retain an international standing, while many academic departments (in polytechnics as well as universities) have found it possible successfully to make their way in the half-commercial world of which the previous government was enamoured. None of that is all that difficult, or in itself offensive. The more serious threat to the research community lies in the hints provided by university entrance figures of the alienation of young people from technical careers and in the certain if anecdotal evidence that more senior people are heading steadily overseas. For Britain as a whole, the risk is that the impending shortage of trained researchers, required as much by industry as by academic institutions, will go unremarked until recovery is virtually impossible.

That is why, on some occasion, there should be a debate in the House of Commons (or somewhere) with much broader terms of reference. The endless argument

about money is not unimportant, but there is a more interesting argument to be had about the means that might be taken, within the debris of the past decade's neglect, to begin reconstructing British science. That will ultimately require a reconstruction of some part of the educational system, some means of giving academic institutions a flexibility they lack at present and a means of persuading young people that science is as much fun as earlier generations have found it. As the chief potential beneficiaries, researchers themselves should be prepared to lend a hand in this cause. □

Hobson's choice?

Germany's new education minister, an easterner, will have a struggle to prove himself.

SINCE the reunification of Germany, it has been clear that places would have to be found in Chancellor Helmut Kohl's cabinet for a sprinkling of easterners, but there is some dismay in academic circles that the education ministry has been one of the first portfolios dealt with in this way. There are several causes for anxiety. First, the education ministry is relatively new, having been carved out of the research ministry as recently as 1972. But in German education, the states or *Länder* are constitutionally dominant, so that a minister needs guile and political skill if he is to exert influence. Yet the new minister, mathematician Rainer Ortleb from the University of Rostock, has virtually no experience of western-style government, and will certainly begin as a pale shadow of his predecessor Jürgen Möllemann who has been promoted to the economics ministry.

It is to Ortleb's credit that he has managed to survive in high office for half a year without falling prey to the investigations of journalists and prosecutors seeking to demonstrate links between eastern politicians and the former East German secret police. Ortleb was a member of the Liberal Democratic Party, which dutifully delivered votes to the old Communist party until 1989. On reunification and the merger of the liberal democrats with West Germany's free democrats (who are the junior partners in the Kohl coalition), Göttele floated to the top, becoming one of the three easterners in the eighteen-strong federal cabinet.

Göttele's more immediate difficulty is Möllemann's legacy. Although something of a gadfly, Möllemann knew how to use his post in education to suggest that his ideas were the next best thing to reality. He struck many adventurous blows for the condition of young researchers and women in German universities, in the process angering the *Länder*, irritating the federal government, but winning the applause of the research community. Möllemann also argued against the search for student numbers at universities in exchange for quality. Göttele deserves a chance to show his mettle, but it is likely to be some time before he emerges from Möllemann's shadow.

Bush asks for 13 per cent extra for science

- Education called top priority
- Initiatives in 'global change', computing

Washington

It is budget time in Washington again, the time of year when the president lets his policy priorities be known through the budget he submits to the US Congress.

President George Bush this week asked Congress to approve a research and development budget totalling \$75,600 million, an increase of \$8,400 million or 13 per cent over research and development (R&D) spending for the current fiscal year, 1991. Basic research would claim \$13,300 million of the R&D total, an increase of 8 per cent, while new funds for applied research and development would jump to almost \$60,000 million, a 13 per cent rise.

White House science adviser D. Allan Bromley declares that "it's a good budget", with special emphasis on science, as well as science education. In the light of the fact that the President's budget overall requests funding increases in the range of 2.6 per cent, it is evident that science has fared well.

However, because the 13 per cent figure for R&D increases does not take inflation into account — currently between 5 per cent and 6 per cent — the real increases are not as substantial as the Administration suggests.

The real outcome of the lengthy US budget process will not be known until sometime next autumn when Congress makes its own decisions about the budget. If history is any guide, legislators will ultimately spend more on biomedical research than the president requests, while cutting back somewhat on defence spending. However, in the light of the techni-

cal success of arms such as the Patriot missile in the Gulf war, high-technology military research may find a receptive audience on Capitol Hill this year.

One of the most contentious issues in US science these days is the disparity between the numbers of individual scientists who are applying for grants and the government's capacity to fund them. Bromley notes that within the National Institutes of Health, the Bush budget asks for a 9 per cent increase in funds for investigator-initiated grants. At the National Science Foundation, which also funds a large number of individual investigators, the proposed increase is for 16 percent to a total of \$2,100 million.

Overall in the Bush budget, however, projects that fall into the category of "big science" have fared best. The Superconducting Super Collider is on course at \$534 million for fiscal year 1992. The National Aeronautics and Space Administration may get a big increase for projects including exploration of the Solar System and new launch systems for the space shuttle. And the Human Genome Project, which Bromley describes as a "vital national asset" is in line for a \$334 million increase for basic research.

This would bring the US investment in deciphering the genome to \$5,000 million.

High performance computing and communications, national security R&D and biotechnology were also singled out as priority areas.

Budget proposals for the principal science agencies and science programmes

see next page

RESEARCH AND DEVELOPMENT SPENDING (\$ MILLION)

	1991 Enacted	1992 Proposed	Dollar change	Per cent change
Basic research				
Doubling the NSF budget	2,316	2,722	+406	+18
Basic biomedical research at NIH	4,634	4,968	+334	+7
Human Genome Project	135	169	+35	+26
Agricultural Research Initiative	73	125	+52	+71
Superconducting Super Collider	243	534	+291	+120
Applied research				
Computing and communications	489	638	+149	+30
Energy R&D	676	903	+227	+34
Advanced manufacturing and materials	1,316	1,310	-6	-
HIV/AIDS	1,152	1,210	+58	+5
Moving Fusion Energy from Science to Engineering	275	337	+62	+23
Aeronautics R&D	482	543	+61	+13
Expanding R&D at the NIST	215	248	+33	+15
Defence R&D	37,738	43,247	+5,464	+14
Space exploration:				
Space transportation infrastructure	4,801	5,517	+716	+15
Space science	1,774	2,141	+367	+21
Mission to Planet Earth	954	1,186	+232	+24
Mission From Planet Earth	2,199	2,470	+271	+12
Biotechnology	3,788	4,107	+319	+8

UK SCIENCE

All change at the top

London

THE Universities Funding Council (UFC) has poached the chairman-elect of the Committee of Vice-Chancellors and Principals (CVCP) to be its next chief executive. Graeme Davies, vice-chancellor of the University of Liverpool, was to have taken over at the CVCP at the end of June, but will now replace Sir Peter Swinnerton-Dyer at the UFC on 1 April.

The universities must wait to see how the balance of power between the new UFC chief executive and his chairman, Lord Chilver, evolves. There have been many reports of friction between Chilver and Swinnerton-Dyer, who chaired the old University Grants Committee before its replacement with the UFC in 1989. Some vice-chancellors fear that Chilver, an advocate of university expansion without a proportionate increase in government funding, will now assume a stronger role. P.A.

Raising the profile

THE funding crisis currently facing British science is to be debated in the House of Commons this week. Jeremy Bray, science and technology spokesman for the opposition Labour party, which requested the debate, says that Labour will attack the "creative accounting" that has allowed the government to claim that the 1991-92 budget for the research councils will keep pace with inflation.

In recent weeks, British science policy has shown some signs of moving from its customary place on the political backburner. The launch of a major policy inquiry by the Royal Society (see *Nature* 349, 183; 17 January 1991) was followed last week by a meeting between Prime Minister John Major and a small group of leading British scientists, organized by Conservative Member of Parliament Sir Ian Lloyd.

Both Downing Street officials and the scientific delegation said they were satisfied with the one and a half hour meeting, although any major policy changes still seem to be off the short-term agenda. P.A.

Polys gain

THE Polytechnics and Colleges Funding Council (PCFC) will spend an extra £7.3 million on research in 1991-92, in addition to its existing spending of about £30 million. The money will be distributed by a new committee, yet to be appointed, to selected polytechnics and colleges with a good research record.

The move follows a PCFC-sponsored report which last year recommended that the PCFC should increase its research spending by £28 million a year.

Peter Aldhous

Budget boost for science

Washington

SCIENCE fares well in this year's US budget proposals — with a 13 per cent increase compared to a 2.6 per cent rise in the budget overall. Here are the details.

National Institutes of Health

President Bush has asked for a record total of nearly \$8,800 million for the NIH, including nearly \$830 million for AIDS research alone. At a stated increase of 5.8 per cent, once inflation is taken into account, the AIDS research budget is holding steady. The current NIH budget is \$8,200 million.

Disputes over the NIH budget have recently centred on whether the institutes can fund 6,000 new grants a year, a figure that advocates of investigator-initiated awards have set as something of a holy grail. Under the Bush budget, even with an increase of 9 per cent for individual grants, it is not likely that NIH can fund 6,000 new and competing grants at the full amount a researcher requests.

Department of Energy

The \$6,400 million R&D budget is up only 4 per cent from 1991, not likely to be enough to cover inflation, but it somehow finds funds to keep the Superconducting Super Collider on track with \$534 million. That is a 120 per cent growth from fiscal year 1991 as the projected \$8,200 million project begins to build up. Support for other high-energy and nuclear physics research is projected at \$1,010 million, a 12 per cent increase.

DOE will get the lion's share of the \$903 million planned for Bush's National Energy Strategy: \$653 million, up 26 per cent from fiscal year 1991. Much of the total will go towards reducing energy consumption and finding alternative fuels, but \$314 million is earmarked for advanced energy technologies, including solar power, superconductivity and R&D on advanced nuclear reactor designs. The fusion programme would be boosted by 23 per cent to \$337 million, part of it allotted for R&D and design work on a burning plasma experiment at Princeton's Plasma Fusion Laboratory and part to continue participation in the collaborative International Thermonuclear Engineering Reactor.

NASA

The \$15,700 million budget, a 13 per cent increase over 1991, includes money for boosting basic space science research, expanding space exploration and improving the reliability of getting into space. To provide for improvements to the Space Shuttle as new technology becomes available, Bush proposes a new \$122 million Assured Shuttle Availability programme;

in tandem, the budget asks for \$350 million, to be split evenly between NASA and the Department of Defense, for development of a new launch system.

To pay for a burgeoning crop of space science studies, including the Mars Observer, the Gamma Ray Observer and the Advanced X-Ray Astrophysics Facility, the budget calls for \$2,100 million, a 21 per cent increase over this year's figures. The Space Station Freedom fares relatively poorly in the budget — only an 8 per cent increase to \$2,200 million — but exploration of the Solar System would draw \$256 million, a jump of 65 per cent, to be shared with the energy and defence departments.

National Science Foundation

President Bush once again reaffirmed his commitment to doubling NSF's budget by 1994 with a 17.5 per cent increase over fiscal year 1991. If the increase is approved by Congress, NSF would be "back on the doubling track" that began in 1987 under the Reagan administration, says NSF acting director, Frederick M. Bernthal.

afforded to education by the Bush administration, and as part of a new FCCSET initiative, funding for education fares well. Support for pre-college science and mathematics education is up 19 per cent to \$253 million, with a 30 per cent increase (\$133 million) at the undergraduate level.

NSF has requested \$23.5 million to begin construction on two identical, but widely separated, detectors that form part of the laser interferometry gravitational wave observatory, and \$16 million towards the first of two 8-metre optical/infrared telescopes, a collaborative research effort to include the United Kingdom and, possibly, Canada.

A request of \$50 million has been made to fund a new instrument initiative that is aimed primarily at individual researchers. Grants of between \$100,000 and \$2 million will be awarded as part of a competitive merit-based programme.

One stipulation is that the federal allocation must be matched equally with funds from non-federal sources.

Science education

One of the toughest issues facing the United States is the dismal state of education in science and mathematics from grammar school right on through college.

MATHEMATICS, SCIENCE, AND ENGINEERING EDUCATION
(\$ MILLION)

	Enacted 1991	Proposed 1992	Dollar change	Per cent change
By educational level:				
Pre-college	515	661	+146	+28
Undergraduate	417	477	+60	+14
Graduate	784	803	+19	+2
Total	1,716	1,941	+225	+13
By agency:				
Defence	416	416	—	—
Education	235	330	+95	+40
Health and Human Services	486	513	+27	+6
NSF	372	456	+84	+23
Other	207	226	+19	+9
Total	1,716	1,941	+225	+13

NSF funding for a high-performance computing and communications initiative, which is coordinated through the Federal Coordinating Council on Science, Engineering and Technology (FCCSET), will total \$213 million, a 26 per cent increase. With contributions from the Defense Advanced Research Projects Agency, DOE, NASA, and other agencies, the computing initiative is budgeted at a total of \$638 million, up 149 per cent from fiscal year 1991.

A second FCCSET initiative, the US Global Change Research Program, is projected for \$1,186 million, a 232 per cent leap, divided among NSF, NASA, DOE, the National Oceanic and Atmospheric Administration and several others.

NSF's share of the programme is \$118.5 million, a 36 per cent increase.

In keeping with the high priority

Bromley has called US science and mathematics education a "scandal", a point that the budget document picks up in less vivid terms. The typical US school year is short at 180 days, compared with Japan, where students are in school 240 days a year, or Europe, where the school year is 200 days.

"Moreover", the budget statement notes, "American students do not work very hard". Although individual states bear the primary responsibility for education in the United States, the Bush budget proposes spending through a range of federal agencies on teacher training, curriculum reform and technical assistance in science and mathematics teaching, for a total expenditure of \$1,941 million, up 13 per cent from this year.

Barbara J. Culliton, Robert Pool & Diane Gershon

ROSAT

How to rescue a satellite

London

JUST a week short of completing an all-sky survey of X-ray sources, the multinational Röntgen X-ray observatory ROSAT span out of control on 25 January with damage to two instruments in its scientific payload. British researchers connected with the project report that power and communications have now been restored, but the satellite has yet to resume its scientific operations. On Monday, project scientists were still undecided about whether to re-start operations using back-up instruments or spend valuable time devising extra safety precautions.

A collaboration between Germany, Britain and the United States, ROSAT was launched last year (see *Nature* 345, 465; 1990) aboard a Delta II rocket from Cape Canaveral, and has so far been very successful. The satellite passes over the ground station at Weilheim, near Munich, for bursts of eight in every 96 minutes over a period of eight hours. For the remaining 16 hours each day, the satellite is effectively on its own.

RADIOACTIVE MATERIALS

New UK road rules

London

BRITISH controls over the road transport of radioactive materials will be overhauled if a private member's bill, now being debated by a House of Commons committee,



becomes law. The bill's sponsor, Conservative Member of Parliament Dudley Fishburn, says the movement on Britain's roads of some 500,000 shipments of radioactive isotopes each year (mostly used in medicine and non-nuclear industries) is effectively unregulated, relying on an outmoded 1948 Act of Parliament.

Strict regulations on the packaging, labelling and handling of radioactive substances will follow if the bill is passed, to bring UK law in line with standards laid down by the International Atomic Energy Agency. Britain agreed to enforce these standards by 1990, but an attempt last year to introduce a law failed, after argument over whether the bill applied to Northern Ireland. Fishburn expects no similar problems this time around. Peter Aldhous

When ROSAT was detected in its first pass over Weilheim at 02:32 GMT on Friday, 25 January after just such a gap, it was clear that it was in difficulties. Onboard batteries were running dangerously low and no telemetry was being received, indicating that something was wrong with the attitude control system — the antenna was not pointing at the ground station and the solar panels were not oriented properly, so that the solar batteries could not recharge. "House-keeping" signals from the distressed spacecraft showed that onboard voltage had dropped below the 1.1-volt safety threshold and was falling towards 0.5 volts — the "point of no return", according to UK project leader Professor Ken Pounds of the University of Leicester.

Engineers from the manufacturer, Dornier, were summoned from their beds at 06:19, when an emergency was declared. Mission Control at the German Space Operations Centre at Oberpfaffenhofen, 25 km from Munich, was able to harness the satellite by 09:12, less than seven hours after the problem was reported. The NASA Deep Space Network (DSN) had by that time been called in to help, and the DSN station at Goldstone, California, began to track the satellite at 10:39 GMT.

Alan Harris, one of the British ROSAT researchers based with their German colleagues at the Max Planck Centre for Extraterrestrial Physics in Garching, north of Munich, says that the onboard computer had lost touch with the attitude control system, possibly due to a software error. As the spacecraft tumbled, its sensitive X-ray detectors pointed directly at the Sun, and were damaged when their emergency 'sunshades' failed to engage in time.

The German X-ray telescope has four instruments in the focal plane, arranged as two redundant pairs. Damage occurred to one of the pair of position-sensitive proportional counters (PSPCs) as sunlight punctured its protective shield, allowing some gas to escape before a safety valve managed to save the duplicate instrument. The first PSPC is now permanently 'blind', but researchers are working on the duplicate which should be "all set to go", says Pounds, after calibration tests. The instruments are a duplicate pair of high-resolution imagers, designed for detailed study of isolated X-ray sources. These were not in use at the time and escaped damage.

The British instrument, the wide-field camera, lost one out of its eight filters, one of a pair used in the all-sky survey. The damaged filter has a back-up, like everything on board ROSAT. "We've lost very little", says Harris.

The cause of the initial failure is still unknown, but Pounds suspects that it

might have been connected with a very large solar flare and an associated surge of protons on the same night. Harris says that there is no firm idea about the cause. Meanwhile, scientists are deciding whether to risk resuming operations or to wait for a time — possibly a period of weeks — for new software protocols to be designed and tested. ROSAT's viewing timetables are scripted to the second, so compromise is unlikely: either it works fully or not at all. The DSN could be used to cover the 16-hour gap as a safety measure, but even with the help of Goldstone and the DSN's Canberra antenna, there would still be a gap of 8 hours that Harris says is "too long for comfort".

Pounds echoes the mood of caution that has descended on Garching. "It's absolutely right that the Germans are taking a cautious approach", he says, in case the problem strikes again. As it was, luck and prompt action saved the DM560-million satellite: just hours later, the batteries would have expired, bringing the mission to an irrevocable close.

Henry Gee

■ Next week (14 February), *Nature* will publish the first X-ray pictures of the Moon, taken by ROSAT.

GENE THERAPY

Cancer trial starts

Washington

THE first clinical trial to use gene therapy to treat cancer began last week at the US National Institutes of Health (NIH) with the treatment of two patients, a 29-year-old woman and a 42-year-old man. The team of NIH researchers, led by Steven Rosenberg of the National Cancer Institute, plans to treat up to 50 terminally ill cancer patients who have malignant melanoma, an advanced form of skin cancer.

This latest trial builds on Rosenberg's earlier gene transfer experiments in which he showed that tumour infiltrating lymphocytes (TILs) — cells with anti-tumour activity — that were isolated from the patient's tumour, marked with a gene for neomycin resistance and readministered to the patient, 'homed' to the tumour site. About half the patients with advanced melanoma that were treated with gene-modified TILs showed some improvement after therapy.

Now, Rosenberg is attempting to use TILs as vehicles to deliver to the tumour site molecules that may enhance their anti-tumour activity. By inserting the gene coding for tumour necrosis factor (TNF) — a potent protein molecule that has been shown to cause the shrinkage of tumours in mice — into TILs, it is hoped to spark localized production of TNF at the tumour site at levels that will cause tumour regression.

Diane Gershon

MRC breach spurs reforms

London

WILHELM Feldberg, a former researcher at the Medical Research Council (MRC)'s National Institute for Medical Research (NIMR) in London, last year breached Britain's 1986 law governing animal experiments, an MRC-appointed independent inquiry has concluded. The MRC accepts the findings, and has already embarked on a series of reforms. But the Home Office, responsible for regulating British animal experiments, rejects the criticism that it 'jumped the gun', in allowing Feldberg to heat the abdomens of anaesthetized rabbits with a lamp bulb, before the MRC had decided to continue funding the work.

Last May, the Scottish animal rights group Advocates for Animals submitted video evidence to the Home Office, showing Feldberg carrying out surgical procedures at NIMR on rabbits which were apparently only lightly anaesthetized (see *Nature* 345, 190; 1990). Feldberg was hoping to understand clinical diabetes by studying the rise in blood sugar level after an animal's body surface is heated.

The inquiry, chaired by Sir Brian Bailey, chairman of a regional British television company and former MRC member, agreed that at least one rabbit suffered unnecessarily. Releasing the report last Monday, Bailey said that, after visiting NIMR and five other MRC research units, he was satisfied that the Feldberg case was an isolated incident, and not "the tip of the iceberg".

The report concludes that, apart from the failure properly to anaesthetize experimental animals, Feldberg broke the law in the summer of 1989, when he continued with his heating experiments after NIMR's named veterinary surgeon had told him to stop. Feldberg's Home Office licence was amended to allow this procedure in September 1989, but the MRC decided not to fund the work in March 1990.

The report criticizes the MRC for deferring this decision, which was "generated by compassion" for a respected, but ageing, researcher, rather than by scientific judgement. The Home Office failed adequately to weigh the benefit of the research against the adverse effects to the experimental animals involved when they amended Feldberg's licence, the report says.

"We think the report has got it wrong", Home Office minister Angela Rumbold said on Monday, pointing out that two of the three referees who reviewed Feldberg's MRC grant proposal had seen some scientific value in the work.

Under the 1986 Act, animal experiments must be carried out in designated laboratories, at which one certificate

holder is responsible for ensuring that regulations are adhered to. NIMR director John Skehel, the certificate holder, is criticized in the report for giving Feldberg permission to work for a further five months after the MRC decided not to fund the project in March 1990. (In the event, the Home Office revoked the licences of Feldberg and his technician that May, once the allegations were made).

Although Skehel failed "technically" to prevent breaches of the law at NIMR, the report says that the fault lies more with the organization of NIMR, where hundreds of researchers are based; the heads of NIMR's 22 research divisions should be made personally responsible for ensuring

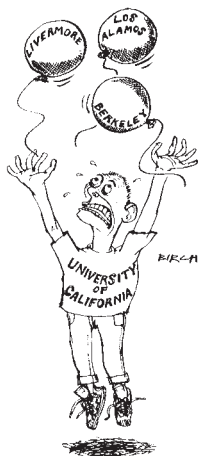
UNIVERSITY OF CALIFORNIA

Stronger management tactics for laboratories

San Francisco

IN response to charges of inadequate management, University of California (UC) President David P. Gardner has unveiled a plan to strengthen the university's ties with the three US Department of Energy (DOE) laboratories it oversees.

The plan for stronger management proposes a 20-member President's Council



on National Laboratories and a major reorganization of Gardner's office, making two senior vice presidents responsible for oversight of laboratory operations. In recent years, critics have claimed that lax management has allowed environmental, security and drug-abuse problems to develop at the laboratories.

The proposal, presented to the UC Board of Regents this month, will also give UC faculty a greater role in monitoring the Lawrence Livermore National Laboratory, Lawrence Berkeley Laboratory and Los Alamos National Laboratory. Last year, the faculty on eight of the nine UC campuses voted to sever the 47-year relationships with the laboratories, arguing that it is fundamentally inappropriate for the university to be involved in the nuclear

that no future breaches occur, and NIMR's named veterinary surgeon should be given greater powers.

MRC secretary Dai Rees said that all the report's recommendations will be implemented — the MRC has already set up a Laboratory Animals Secretariat, to monitor and advise on animal experiments. But there will be no MRC disciplinary action: "We don't see any point in slapping wrists", Rees said.

Nevertheless, the Home Office may still pursue legal action. Its Animal Procedures Committee (APC) meets today (7 February) to consider the findings. The report also asks the Home Office to review each year the licences of researchers over 70 years of age (Feldberg was 89 when the breaches took place).

Peter Aldhous

weapons research conducted at the Livermore and Los Alamos sites. But despite this opposition, the Board of Regents voted in September 1990 to renew five-year contracts with the DOE to manage the laboratories (see *Nature* 347, 322; 1990). Having lost that battle, the faculty has made no formal protest against the new proposal.

UC has tried to improve its relations with the laboratories in recent years, and last year appointed full-time on-site liaison officers at Livermore and Los Alamos, but the new plan is expected to add significantly to UC involvement. The proposed council would be composed of prominent business and government representatives as well as faculty members. It would be charged with reviewing UC management, monitoring the laboratories' operations and policies, assessing new research programmes and offering independent advice to the laboratories' directors.

UC senior vice presidents William R. Frazer and Ronald W. Brady will be assigned to review various research programmes, strategic planning and administrative and environmental policies.

Gardner expects the plan to be implemented as soon as DOE contract negotiations are completed, possibly as early as July. The current contracts, through which UC receives \$12 million a year to manage the 18,000 employees at the three sites, expire in September.

Laboratory officials, who are generally pleased to have access to the support and facilities offered by a closer relationship with UC, say that enhanced interactions make sense now as weapons research is scaled back. The new proposal may ensure a long-term role for the university as laboratories' scientific goals expand in areas such as alternative energy sources, global environmental changes and biomedical sciences. Elizabeth Schaefer

GULF OIL SPILL

Big test for bioremediation

London

TIME, and the action of oil-feeding bacteria, will be the only healers for the Western Gulf coast ecosystem, now threatened by the largest oil slick in history, most marine pollution experts agree. But the slow recovery may be accelerated by the relatively new techniques of bioremediation — adding nutrients or micro-organisms to supplement the natural process of biodegradation.

Oil spills have traditionally been tackled using dispersant chemicals, or by removing the oil physically from beaches and the sea surface. But chemicals must be applied to spilled crude oil quickly, before the volatile components evaporate and

new slick “overwhelms the spill-response capability of the Gulf region”, Golob says. Left alone, the oil will linger for years.

Bioremediation efforts should be less hampered by the shortage of manpower. Beaches and floating oil can be sprayed periodically, and the bacteria left to carry out their task. But the techniques are relatively new, and nothing has been attempted on this scale before.

After the *Valdez* spill, Exxon sprayed several hundred kilometres of the Alaskan coastline with fertilizing nutrients, to stimulate the natural populations of bacteria. Oil degraded several times more quickly on these beaches than on untreated sites, says Dick Lessard, Exxon's oil spill technology co-ordinator.

Ron Atlas, from the University of Louisville, believes spraying beaches with nutrients can also help in the Gulf. In theory, the warmer conditions on Gulf beaches will help, if the *Valdez* techniques are repeated, as the oil-feeding bacteria should be metabolically more active. Despite the larger volume of oil in the Gulf spill, Atlas says that a smaller proportion of the total is likely to reach the shore than in Prince William Sound, which is littered with small islands. Experience with the world's largest previous oil spill, the 3-4 million barrel Ixtoc well blow-out in the Gulf of Mexico in 1979, shows that oil staying offshore poses a smaller ecological threat, Atlas says.

Bioremediation has also been tried on slicks at sea, in the Gulf of Mexico, on a spill from the tanker *Mega Borg* in June last year. A US Coast Guard vessel sprayed a cocktail of oil-feeding bacteria and fertilizer, together with a biological catalyst. Although the oil disappeared from the treated areas within 16 hours, it is not known whether the treatment, or simply wave action, was responsible. Atlas says there have been no experiments that conclusively prove that bioremediation can work on offshore slicks.

Nevertheless, the Texas company Alpha Environmental, which supplied the bacteria for the *Mega Borg* clean-up, and its UK sister Alpha Biological Treatment Services are standing by to move two and a half tonnes of their dried bacterial cocktail and 30 oil-drum-sized fermenters to the Gulf. George Hubbard, joint managing director of the UK company, maintains that it will be possible to grow enough bacteria to have a “considerable” impact on the slick, both at sea and on the shore, but adds “it's going to take months”. The Alpha cocktail contains some 50 naturally occurring marine bacteria, collected from around the world.

Lessard and Atlas believe it will not be necessary to add new bacteria in the Gulf,

along with fertilizer, because previous smaller spills should have nurtured a healthy population of oil-degrading bacteria. But Atlas sees no harm in Alpha's plans, provided that any potentially pathogenic bacteria are kept well away from the desalination plants.

Tony Knights, from the University of Wales at Swansea, who worked on marine pollution in the Gulf in the 1970s and 1980s, suggests a low-technology alternative approach to attack the spill while still at sea: the slick could be sprayed with cement powder to bind the oil and make it sink *en masse*, away from environmentally sensitive areas. But other experts are sceptical of this method. Madeline McDonagh, from the UK Department of Trade and Industry's Warren Spring Laboratory, says sinking the oil will delay its breakdown, and may create problems in the long term, as the oil seeps back to the surface.

Whatever the scope for fighting the existing slick, the biggest fear is that further huge spills are possible, into an almost enclosed sea that averages only 30 metres deep.

Peter Aldhous

INDIA

Impact of Gulf War

New Delhi

THE Gulf War will have a negligible impact on India's environment and weather according to an assessment by the country's science agencies, but the government has warned of “possible serious long-term adverse effects” on the economy.

Even if the oil slick in the Gulf enters the Arabian Sea, it will not influence the monsoon due to set in on 1 June, says the Department of Science and Technology (DST).

The monsoon depends more on what happens over the Tibetan plateau and the Bay of Bengal than on the Arabian Sea, says Vasant Gowariker, secretary of the department. Conditions in the Arabian Sea are not even included in the model that the DST has used successfully to forecast monsoon behaviour in the past three years. The Arabian Sea is very rough at this time of year and any slick entering it will break up and evaporate leaving tar balls, which Indian coastguard vessels are standing by to scoop up.

A joint study by defence and civilian scientists says pollution caused by fires at oil wells will have no adverse effect on agriculture or the weather but warns that the economy will be hit by damage to Iraqi oil wells — 25 per cent of India's oil imports come from Iraq. The government is responding to the threatened energy crisis by enforcing conservation measures, switching to coal wherever possible, and encouraging alternative energy projects such as producing electricity from animal and tidal power.

K. S. Jayaraman



Desalination plants threatened by slick.

the oil weathers to a tar-like consistency. With the main source of the slick (the offshore tanker loading station at Mina al-Ahmadi in Kuwait, where oil was released by Iraqi forces) in the middle of a war zone, this has proved impossible. Chemical treatment as the oil drifts down the Saudi Arabian coast would in any case be unwise, because dispersed oil droplets would sink, and could then enter the inlet pipes of Saudi desalination plants.

Given the huge volume of oil (probably more than 10 million barrels) and the limited time available, the deployment of oil-containing booms and sea-surface-skimming equipment will be limited to protecting economically important installations, such as the desalination plants. Richard Golob, publisher of *Golob's Oil Pollution Bulletin*, says the slick's scale will also make any physical clean-up after the oil comes ashore difficult. More than 10,000 people were drafted in to clean beaches in Prince William Sound, Alaska, after the 1989 *Exxon Valdez* disaster, where 250,000 barrels were spilt. The

Restructuring gets underway

Munich

In a move that enraged researchers in the former East Germany, but created hope for change in an inflexible research structure in the west, the science council Wissenschaftsrat last week recommended drastic staff cuts and an entirely new structure at three research institutes in East Berlin that used to be attached to the East German Academy of Sciences.

The institutes are the Central Institutes for Molecular Biology (ZIMB), for Cancer Research and for Heart and Circulatory Research, which share a campus with two hospitals in Berlin-Buch.

At the same time, the council passed a milder judgment on the (much smaller) former academy institute for high-energy physics at Zeuthen, calling for a majority of the personnel to be maintained and the institute to be integrated into a related institute in the west.

The recommendations for Berlin-Buch and Zeuthen were among the first of many recommendations expected in next few months would put a radically new face on science in eastern Germany and improve the status of German science as a whole. But the council warned that to work its initiatives have to be backed up with money from both Bonn and the *Länder* (states).

NIH in Berlin?

If followed, the recommendation for Berlin-Buch could give Germany its first truly interdisciplinary research institution combining basic biology research with clinical research in the tradition of the US National Institutes of Health (NIH). Wissenschaftsrat called for at least half of the positions to be of limited duration and half of the money to come from research grants, in contrast to the inflexible staff and funding structure of most western German research institutions.

Insiders say that the proposal has only a small chance of succeeding in its present form. Wissenschaftsrat calls for the hospitals in Berlin-Buch to provide 100 beds all the time for patients of particular interest to research. Such an arrangement might have worked in the former East Germany, but as hospital administrators and doctors in the west see each bed as a source of income, it is unlikely to work in the united Germany.

At the same time, researchers at ZIMB were enraged to discover that Wissenschaftsrat had recommended reducing personnel there disproportionately more than at the other two institutes at Berlin-Buch.

Whereas the overall number of researchers would drop from about 1,500 to 500 or 600 if the recommendation is followed, ZIMB director Günter Pasternak calcu-

lated that ZIMB might lose as many as seven-eighths of its 650 permanent staff members.

Although researchers at ZIMB had expected the council to recommend cuts affecting as much as even 50 per cent of the staff, no one was prepared for the scope of the reductions actually proposed. Institute researcher Tom Rapoport says it would be a cruel blow to reduce the staff by so much at one time, not least because the level and duration of unemployment benefits are far from certain. Wissenschaftsrat chairman Dieter Simon says that the fate of the researchers was not part of the council's agenda. "We just made structural recommendations".

Nevertheless, ZIMB researchers familiar with the text of the recommendations say that politics rather than science was the main criterion of the council. They say that while ZIMB researchers received many positive reviews, groups there did not fare as well in the recommendations as groups in the other two institutes, whose focus is much closer to the purpose of the proposed new centre.

Good review for Zeuthen

Like Berlin-Buch, the institute at Zeuthen is also considered to be one of the best that East Germany had to offer. Groups from Zeuthen participated in high-energy physics experiments at state-of-the-art research centres such as DESY (Deutsches Elektronen Synchrotron) in Hamburg and CERN in Geneva as well as on cosmic-ray projects at Lake Baikal in the Soviet Union.

Because it did not want to witness the dismantling of an institute that functions well and might play a role in raising the standards of physics in eastern Germany, the council recommended that the Zeuthen institute be preserved, and linked to DESY as some sort of satellite.

The leadership at DESY has mixed feelings about absorbing the Zeuthen institute. Although DESY's directors are happy that an institution with such a good reputation got good marks from Wissenschaftsrat, the head of the DESY research division, Paul Söding, says that it will be a "difficult question" whether the institute at Zeuthen can be absorbed.

It would be unprecedented for DESY, which builds and maintains particle accelerators, to absorb one of its users, one with virtually no equipment of its own. In addition, there is no money in the current DESY budget (financed primarily by the Bonn Research Ministry) for such an acquisition. Nevertheless, Söding says that DESY will offer any help it can to its new potential colleagues while debating its course.

Steven Dickman

Politics rears its not-so-ugly head

THE Wissenschaftsrat recommendations for the future of the institutes at Berlin-Buch and Zeuthen show that research quality is not the only criterion being considered in reforming German science.

As Wissenschaftsrat members assess the level of research in an eastern German institute or department, they must attempt to integrate any new structure they recommend into a densely populated network of research institutions in the west.

The following factors come into play: How easily can research be integrated into existing structures, such as universities or national laboratories (*Grossforschungseinrichtungen*)? Are there any institutes in the west doing this type of research? How does the institution fit in with the local research infrastructure?

A comparison of the recommendations for the institutes at Berlin-Buch and at Zeuthen (see left) shows how important these factors can be. Reading between the lines, some other reasons are apparent: although only just outside Berlin, unlike the institutes at Berlin-Buch, the Zeuthen institute is in a *Land*, Brandenburg, where there is virtually no basic research and no universities. Berlin, by contrast, is overflowing with research institutions.

Zeuthen was also saved by its international contacts, which extended west as well as east, and by the relatively large percentage of researchers judged to be doing high-level research.

Furthermore, because much of the work done at Zeuthen itself was theoretical, consisting of the analysis of experimental results obtained at DESY, CERN and elsewhere, the institute was not hurt as badly as others by the lack of equipment. Finally, Zeuthen profited by the sheer numbers involved: it is easier to save a majority of 220 researchers than it is to save a majority of 1,500 researchers.

Steven Dickman

Simon stays

THE science council Wissenschaftsrat, wanting a steady hand on the rudder in the difficult process of reunifying German science, has re-elected Dieter Simon for at least one more year as its chairman. Simon, may serve up to three more years in the post. He became chairman in 1989.

Normally, Simon would have had to step down not only from the chairmanship but from the council itself because his two three-year terms as a member had expired. But German President Richard von Weizsäcker, who oversees nominations to the council, urged members of the council's scientific commission to waive the six-year rule in Simon's case. The motion was approved unanimously. SD

GLOBAL WARMING

Meeting set fair for squalls

Chantilly, Virginia

THE United States managed to provide an unseasonably warm February day to herald the first meeting of the United Nations Intergovernmental Negotiating Committee (INC) for a Framework Convention on Climate Change, but inside the conference hall the atmosphere began to turn chilly as soon as the welcoming messages had been delivered. The ten-day meeting marks the start of what is supposed to be an eighteen-month period of international negotiations leading to the presentation, at the June 1992 meeting of the UN Conference on Environment and Development in Rio de Janeiro, of a treaty that will commit its signatories to industrial and economic policies aimed at curbing the threat of global warming (see *Nature* 349, 358; 1991). But the proceedings of the first session of the INC last Monday (4 February) indicated that 18 months could be used up in agreeing on negotiating procedures, without science or economics being mentioned.

After a statement from UN Secretary-General Javier Perez de Cuellar adverting to the "formidable task" ahead of the delegates was read out, the remaining inaugural speakers went through the now familiar issues of scientific uncertainty over the present extent, if any, of anthropogenic global warming, of environmental activism versus economic conservation, and of the need for developed countries to pay, in one way or another, the less developed world not to use old-fashioned polluting technologies as they

attempt to build their economic strength.

Representing the host country was the chairman of the US President's Council on Environmental Quality, Michael Deland, who at the outset stated the US desiderata of achieving "economic growth with responsible stewardship" of the Earth's resources. His generally cautious speech, which emphasized the importance of recognising "financial and technological realities" in working towards an international treaty, was counterbalanced by more combative remarks from Mustafa Tolba, head of the UN Environment Programme. Tolba declared that there was increasing scientific evidence to suggest that "global warming may have begun," and that without action a temperature rise of two to five degrees centigrade during the next century was likely.

Tolba also suggested that disagreements between different economic forecasts of the costs of CO₂ reduction programmes and of the potential costs saved by boosting energy efficiency and maintaining the Earth's climate in its present state were largely due to different assumptions rather than profound disputes over methodology, and that with reasonable cooperation it would be possible for all parties to agree on ways to quantify the costs of global warming and of fighting it.

But in the end it is likely to be fundamental political differences rather than scientific or economic uncertainty that dictates the pace of the INC negotiating process.

David Lindley

NANOTECHNOLOGY

Miti aims at small research

Kyoto

LAST week saw an unusual little gathering of biologists, physicists, and chemists in Kyoto to discuss atomic-level design of functional structures. The workshop, which formed part of the 10th Science and Technology Forum organized by the Science and Technology Agency (STA) is just one example of blossoming interest in Japan in nanotechnology, a field which with the backing of the powerful Ministry of International Trade and Industry (MITI) seems destined to become Japan's next priority target for industrial research.

In recent years several nanotechnology projects have been launched under STA's imaginative Exploratory Research for Advanced Technology (ERATO) programme which provides teams of young researchers from academia and industry with annual funds of 2-3 million dollars over a period of five years to develop the seeds of new technology.

Now MITI's Agency of Industrial Science and Technology (AIST) is getting in on the act with the launch later this year of a project under the agency's "basic technologies for future industries" (Jiseidai) programme. The project, which Sakaki will participate in, will develop quantum dot and quantum wire devices with a MITI budget of about 5,000 million yen (\$40 million) over ten years.

Private companies, which typically number 5-8 for Jiseidai projects, will be selected this summer and will include some of Japan's major electronic manufacturers.

MITI sees nanotechnology as opening the way to development of the next generation of gigabit electronic devices early in the next century. But the ministry is also keen to support nanotechnology research because it involves a lot of pure science and a large-scale project might help diffuse Western criticism of Japan's perceived lack of contribution to basic research.

David Swinbanks

NUCLEAR POWER

Opposition melts away

Tokyo

JAPAN's nuclear power industry breathed a sigh of relief on 3 February when a pro-nuclear candidate was elected governor of Aomori Prefecture on the northern tip of the main island of Japan. The local election was seen as a test of public support for ambitious plans by the central government and industry to establish a huge nuclear fuel complex in Aomori for the enrichment of uranium, reprocessing of spent fuel and storage of nuclear waste (see *Nature* 345, 285; 1990).

The Aomori complex, which is already under construction in Rokkasho village, has become the focus of attention of a growing anti-nuclear power movement in Japan which has recruited support from many of Japan's housewives following the Chernobyl disaster in 1986. A Rokkasho government official who advocated freezing the project was recently elected mayor of the village but then showed little opposition to the complex. And the anti-nuclear movement was pinning its last hopes on the governor's election.

But the incumbent governor, Masaya Kitamura, who has long supported the project, won by a surprisingly comfortable margin against a candidate who advocated a freeze and another proposing total abandonment.

David Swinbanks

RIBOZYME TECHNOLOGY

Cech wins first patent

New York

MORE than four years after the first application for ribozyme technology was filed, the US Patent Office has granted the first patent (Number 4,987,071) to Thomas Cech and the University of Colorado.

Ribozymes, RNA catalysts that can specifically cleave other RNA molecules, are likely to have important pharmaceutical and biotechnological applications because of their ability to destroy RNA molecules that code for individual proteins.

Thomas Mann, chairman of United States Biochemical Corporation (USB), which holds an exclusive licence to Cech's work, says the patent covers "a broad range of catalytic RNA molecules, methods of using them and methods of synthesizing [them]". This includes RNA enzymes that cleave RNA at any site except a specific four-nucleotide sequence, CUCU, for which USB already has a commercially available ribozyme product.

The new patent contains 63 separate claims, 10 fewer than when the application was first filed, and has been extensively revised since that time. Other applications are still under consideration, and according to Richard Warburg, a patent lawyer with Fish and Richardson in Boston, a second patent may be awarded to Cech later this year.

Kevin Davies

Kuwait after the invasion

SIR—I have been teaching at the faculty of medicine, Kuwait University, since September 1978, when the university began its first intake of preclinical students. Since 1984, I have been associate professor of biochemistry.

I was in Kuwait at the time of the Iraqi invasion on 2 August and have only recently returned home following the decision of the Iraqi authorities to release all Western hostages. Although unable to leave Kuwait or Iraq, I was as a Canadian citizen able to travel relatively freely within Kuwait City and was thus able to visit the medical school on several occasions following the invasion. While I am sure that your readers have read the newspaper reports concerning the atrocities committed by the Iraqis in Kuwait (for many of which I have myself listened to first-hand accounts), I believe your readers, including among their number so many of the world's academic and scientific community, should know of the extensive looting and virtual extinction of the medical school.

During the period following the invasion, I was able to witness at first hand the wholesale removal of facilities and equipment. Among items removed were the entire library and major equipment as well as chemicals. The laboratories had been for the most part ransacked, with broken bottles and their contents littering the floors.

During conversations with former students (both Kuwaiti and other Arab nationals), I was informed that they had witnessed professors from Iraqi universities squabbling with each other over what items of equipment were to be despatched to their individual departments and laboratories. This behaviour goes far beyond the minimal compliance with the military authorities that might have been excusable on the grounds of personal danger.

At both undergraduate and graduate levels, the students have been devastated by the dismantling of the university. Undergraduate students have spent up to six years of study at the university and for many the prospects of their completing their studies after so many years of effort look very dim. For the postgraduate students, vital equipment, animals and results relating to their thesis work have been lost or destroyed, and the same of course applies to faculty members who have seen the irreversible loss of many years' work. In a cynical move, the Iraqi authorities have offered places to students to continue their education at Iraqi universities. Not only does this mean leaving their country, homes and families, but in addition, the offer is conditional on Kuwaitis renouncing their citizenship. For

both Kuwaitis and non-Kuwaitis, there is every likelihood that after transferring to Iraq, they would be called upon to serve in the Iraqi armed forces.

After many years of dedicated effort by Kuwaiti and expatriate staff, the medical school had reached standards recognized by many internationally respected academics and scientists who had visited the faculty in various capacities including as external examiners. The destruction of the medical school is a tragedy not only for Kuwait but for the whole of the region. A time will come when the international academic and scientific community should consider what action they might take to ensure that those responsible for these deeds are compelled to assist in the restoration of the Kuwait University Medical School.

GARY BURNS

*Freie Universität Berlin,
Institut für Virologie,
Nordufer 20, 1000 Berlin 65, Germany*

Radiation risks

SIR—In assessing the decision to reduce the limits for annual radiation doses to radiation workers (*Nature* 348, 274; 1990), it is important to distinguish between regulatory wisdom, however developed, and scientific substance. Although the decision is based on some scientific evidence, it also relies on the linear hypothesis which assumes that the biological effects of high levels of radiation can be extrapolated linearly to zero dose.

While this hypothesis is convenient, its predictions correlate poorly with epidemiological data on the effect of exposure to low levels of radiation (that is, below 100 millisieverts, mSv). There is no consistent evidence of any hazard due to radiation at these levels^{1,2}; indeed, some data suggest that those irradiated have reduced risks³. For instance, people living in areas with high background radiation tend to have a lower incidence of cancer⁴, and the average lifetime of the Japanese bomb-survivors now exceeds that of other Japanese⁵. The apparently increased risk of leukaemia in children of workers at the Sellafield nuclear reprocessing plant has not been corroborated, nor other explanations excluded (see *Nature* 343, 679; 1990). But the International Commission on Radiological Protection recently reduced the permissible limits for the public to 1 mSv, well below many natural background levels, and has now further reduced the limits for radiation workers below the levels at which any hazard has been demonstrated.

While accepting the desire to maximize the safety of the public, including those with occupational exposure to radiation,

we question whether the long-term effects of these unrealistic radiation limits will be beneficial. The overall costs arising from compliance with the standards and from distortion of the roles of nuclear industries may be enormous.

F. C. G. SOUTHON
S. C. GANDEVIA

*Institute of Neurological Sciences,
Prince Henry Hospital,
PO Box 233, Matraville, NSW 2036,
Australia*

1. Webster, E. W. *Am. J. Radiol.* **137**, 647 (1981).
2. Trott, K. R. & Streffer, C. in *Radiation Exposure and Occupational Risks*, 61–74 (eds. Scherer, E. et al. Springer, Berlin, 1990).
3. Mettler, F. A. & Moseley, R. D. Jr *Medical Effects of Ionizing Radiation*, 90 (Grune & Stratton, London 1985).
4. Nambi, K. S. V. & Soman, S. D. *Health Phys.* **59**, 339 (1990).
5. Cameron, J. *Am. Ass. of Physicists in Med. Newsletter* **15**, 3 (1990).

Romania rebuilds

SIR—The leading article "Will recession become slump?" (*Nature* 348, 96; 1990) flatly states that "prices have also increased in Romania and Bulgaria, so far without much compensating freedom", implying that nothing in Romania has changed.

In Romania, civil liberties have taken major steps forward: the springing up of numerous uncensored publications, political discussion in the streets and public places, unfettered and open meetings of individuals and organizations, unhindered meetings between foreigners and Romanians and open party politics.

In response to the December Revolution and the freedoms arising from it, a group of eminent ecologists, engineers, doctors, lawyers and writers founded the first private university in Romania, the Ecological University of Bucharest. Classes began last fall, only nine months after the revolution. The faculty had planned for 1,200 students but the current enrolment is 1,700. There are six full schools: natural sciences, engineering, medicine, law, physical education and moral ecology.

Like many educational institutions in Romania, the university needs current books and scientific equipment. Sampling and analytical equipment media (gas chromatographs, HPLCs, atomic absorption spectrophotometers with furnace, and the rest) and methodological manuals for environmental sampling and analysis are especially needed.

I should like to hear from anyone who can donate books and/or equipment.

DEBORAH WALLACE

*CBNS, Queens College,
Flushing, New York 11367, USA*

■ The aspirations of the Hungarian-speaking minority of Transylvania and the future of the Hungarian-language minority there are apparently not so cheerful. Editor *Nature*. □

How to slow global warming

David G. Victor

A General Agreement on Climate Change would accommodate the diverse and changing interests of nations as they seek to slow global warming alongside other socioeconomic objectives.

Is it possible to structure an effective climate agreement that gains wide acceptance by nations and is also flexible enough to accommodate new scientific findings and political interests as they arise? Many present conceptions of a climate agreement appear unable to meet these criteria. An alternative would be to structure an open-ended agreement modelled on the iterative negotiation process of the General Agreement on Tariffs and Trade (GATT). A General Agreement on Climate Change (GACC) would consist of core agreements on allowable national contributions to global warming over time and would enable a wide range of other agreements on technology transfer, funding mechanisms and other issues as needed to accommodate the interests of nations.

Although the GATT process is in disarray, the model should be useful to nations as they initiate negotiations for a climate agreement. Indeed, some of the failures of GATT — such as lack of adequate attention to developing countries — form important lessons for GACC. Overall, GACC would begin as a limited agreement and evolve over time. Because of its modest early intentions, it should be possible to negotiate the structure for an adaptive regime such as GACC by 1992. The first round of substantive negotiations could begin at the UN Conference on Environment and Development.

Debate

The theory that the Earth is warming because of an enhanced greenhouse effect now commands wide support among scientists¹. Important debates remain; notably, there are cogent arguments regarding the stringency and urgency of greenhouse policies — if any — which are needed. But the threat of global warming may warrant policy action long before fundamental uncertainties are answered. Indeed, much of the present debate over climate policy acknowledges that the political schedule to slow warming may have to proceed more rapidly than the advance in scientific understanding.

Because of the extremely complex nature of the climate issue, many students of international affairs are pessimistic about the near-term prospects for an agreement to control climate change². Global policy action to control the climate will confront the classic problem of managing the commons³, with the costs

and benefits of collective action to slow global warming widely dispersed across political boundaries. Concentrated interests aligned against the diffuse beneficiaries of climate control will further block action⁴.

Although it may be in the collective interest of the community of nations to slow global warming to some degree, it will be difficult to foster collective action^{5,6}. Cooperation will be easier if there is a leader, but the largest contributors to global warming (the Soviet Union and the United States, which account for 38 per cent of annual fossil-fuel carbon dioxide emissions)⁷ are moving slowly.

Despite the factors working against collective action on the climate issue, numerous countries are unilaterally moving to controls on emissions of greenhouse gases⁸. At least 22 per cent (roughly equal to the US share) of current fossil-fuel emissions of carbon dioxide is now subject to such unilateral controls (computed from the list of countries in ref. 8 and the emissions reported in ref. 7).

Diverse interest groups are lobbying for swift policy action, with environmental organizations adding pressure by publicizing global warming as a leading environmental threat. Scientists in the public eye have stressed the extreme risks of unchecked warming, and climate surprises such as the hot summer of 1988 intensify international support for policy responses. Nations with extreme interest in the deleterious effects of climate change have raised the prominence of the climate issue in fora such as the United Nations⁹.

Further, the 1992 UN conference on the environment and development (UNCED) may help link global environmental policy to a longstanding interest in realigning North-South inequity. States from the South — historically not the largest contributors to global warming¹⁰ but probably large emitters of greenhouse gases in the future^{11,12} (and already the largest sources of carbon dioxide from deforestation¹³) — may have strong additional incentives to cooperate with international greenhouse policy.

Thus while a climate agreement will be very difficult to attain, the international community appears to be rapidly moving towards that goal. The path to a formal agreement starts in Washington this month and many hope that an initial 'framework' agreement will be signed before or during the UNCED. Such an

agreement would be the first step towards formal greenhouse norms and institutions. Over the next months, decisions will be made that could have sweeping impact on the social and economic processes that contribute to climate change. Because the features of a climate agreement may persist for decades, it is essential to study the alternatives closely.

Criteria

How might we structure such an agreement? What criteria must such an agreement meet? This discussion centres on these fundamental questions, developing two essential criteria — political acceptability and long-term flexibility — for assessing rival approaches. Five alternative structures will be considered, and an additional structure — one based on the iterative, flexible negotiations of the GATT — will be proposed.

The focus of this analysis is not how much abatement of greenhouse gases is necessary or desirable. There are no robust criteria capable of providing an answer. That issue will be determined by a bargaining process among interests, arising from compromises on a multiplicity of criteria related to quantity, timing, cost and burden-sharing of abatement.

Throughout this discussion, the level of analysis will be the international system, and it is assumed that the parties to an international agreement will be nation-states. Agreements that might be formally codified between these states are considered. The salient issue is: what structure of a political agreement to slow global warming will allow the necessary bargains among states? For the sake of brevity, structural questions are artificially divorced from other areas of concern.

The goal of efforts to slow climate change is an effective climate regime. Effectiveness will be judged in terms of ability to alter the behaviour of nations (and in turn their constituents) towards smaller contributions to global warming. An effective agreement will be one which gains (and uses) the greatest leverage over activities which contribute to global warming. The strength of that leverage will vary with the extent to which two criteria are met. An agreement structure which best satisfies the criteria will maximize the number of participating nations as well as the extent to which these nations will make sacrifices to slow global warming.

The criterion of political acceptability rests on at least two principles. First, any global agreement must recognize the nation-state unit of international organization and the highly dispersed, interstate nature of the long-term sources and effects of climate change. Consequently, a climate agreement must allow the wide range of national governance styles to prevail while, at the same time, controlling the commons. It is unlikely that Nigeria, Indonesia, Sweden and the United States will agree to use the same domestic policy instruments to reduce emissions.

Equity

The second principle of a politically acceptable agreement is the explicit recognition of equity. A recent review of international environmental agreements highlighted equity as one of the main features of successful agreements¹⁴. Many authors have appropriately focused on the role that a climate agreement must play in addressing North-South inequities^{12,15}. Because the greatest leverage over future greenhouse emissions demands that developing countries participate, a climate agreement must presumably include the enticement of measures to reduce North-South inequities. At minimum, an agreement that is perceived to exacerbate existing inequities would presumably not be acceptable.

Yet a broader conception of equity — one that centres on equitable distribution of costs (and benefits) across all groups, including North and South — will pose the most difficult challenge. Because stringent climate policy will have considerable effects on virtually every aspect of industrial and agricultural societies, it is reasonable to expect that the implications for welfare will also vary considerably among nations. Thus, unlike many other environmental issues which have been relatively peripheral, international cooperation to slow global warming must explicitly link environmental policies to other incentives and penalties to equalize burden-sharing.

In principle, equity in burden-sharing can be adjusted with income transfers; in practice, however, this will be difficult because the needed transfers may amount to tens of billions of dollars. Equity concerns are also addressed by adjusting the national requirements for abatement so that the wealthy and those with strong desire to slow global warming will bear a greater burden. Merely varying the levels of national abatement, however, sacrifices leverage over present and future greenhouse emissions (and thus global warming). Burden-sharing arguments will presumably have to be addressed through a wider range of concessions and incentives and not just financial mechanisms and differential abatement.

The criterion of long-term flexibility has

received much less attention than the aforementioned political criteria. Because the emissions of greenhouse gases are inextricably linked to industrial and agricultural practices at the core of modern societies, a greenhouse agreement must be flexible enough to accommodate social changes. An agreement must adapt to periods of crisis (for example, war), opportunity (for example, waves of environmentalism) and other factors which will affect the extent to which nations want to control climate change *vis-à-vis* other contending objectives. Indeed, reviews of other cooperative arrangements to manage other international issues stress the need for adaptability^{16,17}.

Typical greenhouse projections extend to the year 2050 or 2100; even low projections of greenhouse gas concentrations show that global warming will be a significant concern well into the next century¹⁸. Looking back over the last century of US history, for example, there have been two severe depressions, three extended periods of rapid economic growth, two large wars, and massive changes in the relationship between the government (federal, state and local), industry and the citizenry. All of these factors would have affected the character of a greenhouse agreement had one been in place.

The same dynamism will presumably continue in the future. The terms of the global-warming debate cast today will change even within the decade. Similarly, the technological context within which the effects of global warming are felt will also change¹⁹. A greenhouse agreement must be flexible enough to allow adjustments to the terms of agreement so that the objective of slowing global warming can be achieved alongside other changing socioeconomic objectives and conditions.

How well do the various agreements considered here fare when these criteria are applied? Five different classes of proposed structures for a climate agreement will be presented: (1) command-and-control, (2) Law of the Atmosphere, (3) greenhouse taxes, (4) tradeable permits, and (5) convention-protocol.

The five strategies are conceived of separately for analytic convenience and each one is analysed as if it were the primary means of global greenhouse control. Clearly, however, there are numerous possible hybrid agreements. For example, signatories to a convention-protocol agreement — which requires an implementation strategy at the national level — could use command-and-control, greenhouse taxes and/or tradeable permits as domestic policies. Implementation strategies (command-and-control, taxes and tradeable permits) are distinct from negotiation frameworks (Law of the Atmosphere and convention-protocol). But the two categories are simultaneously considered here because the implementa-

tion of an international greenhouse agreement will clearly affect the process of negotiation (and, by extension, the ability of the agreement to meet the criteria).

The first approach is broadly known as command-and-control. Under this system, a greenhouse authority would instruct nations (and perhaps even individuals) regarding which industrial and agricultural practices to use or how to alter existing practices to lower greenhouse gas emissions. For example, emitters might be instructed to install certain energy-efficient technologies or adopt certain farming procedures that reduce emissions of greenhouse gases.

Variants on command-and-control have been a typical form of environmental regulation until the present time. There have been some notable successes under this system, but many believe that excessive costs arise from managing environmental problems by directive rather than by economic incentives. Furthermore, the successes have been most marked in the case of bans (for example, DDT, PCBs) or similarly strong policy actions.

Few aspects of climate change are well suited for strong policy actions such as bans. Although some studies claim that substantially lower energy and greenhouse paths are possible²⁰, the consensus is that there are probably no acceptable near-term solutions to climate change.

The main weakness of a command-and-control approach for an international agreement is that management by decree provides little opportunity for nations to choose the strategies which best suit their constituents. There is no doubt that if Asian rice or American electricity were produced differently, significantly lower emissions of greenhouse gases could result. And if implemented perfectly, changes in technology or practice could lower emissions at very small or even negative cost.

But sovereign nations will object to instruction regarding their choice of strategies (and the costs of implementing them). Many may choose to apply command-and-control strategies domestically in order to meet the obligations of an international treaty, but those that do not will certainly resist a global command-and-control structure.

Acceptability

Therefore, command-and-control appears to fail the criterion of political acceptability on the principle that it does not allow for diversity of national styles of governance. In theory, however, command-and-control is flexible over time since decrees could be altered to match changing social, economic and technological contexts.

A second approach involves the negotiation of a Law of the Atmosphere (LOA) or other similar comprehensive

agreements (reviewed in ref. 21, along with other international legal approaches). A Law of the Atmosphere would contain rules of access for the common resource of the global atmosphere. Like the Law of the Sea (LOS), it would cut across a number of issues in a comprehensive way. In addition to climate change, a Law of the Atmosphere could accommodate other global and regional atmospheric issues.

Most importantly, a Law of the Atmosphere could directly couple financial mechanisms to the use of the atmospheric common resource. Because the package would be comprehensive, in theory it could accommodate the wide range of equity concerns and could allow national styles of governance to prevail domestically provided that the terms of the international agreement are met.

But the main problem with the LOA approach is its flexibility over time. It will take a long time to reach an accord on such sweeping, varied measures, and it will be similarly time-consuming to modify the agreement as needed in the future. Although a comprehensive agreement such as LOA may need fewer modifications than other possible agreements, those needed changes that do emerge are likely to accumulate and eventually undermine the LOA.

Even if a treaty such as the LOA were negotiated, it would be highly susceptible to catastrophic failure caused by defection of one or more major signatories. In the past, several ambitious, comprehensive treaties have followed this frustrating pattern. The 1946 Baruch plan to transfer control of nuclear weapons to the United Nations, the 1948 International Trade Organization (ITO) to liberalize comprehensively world trade (and other economic practices), and the 1982 Law of the Sea all failed to gain wide acceptance because one or more major parties defected at the last moment.

In each of these cases, the comprehensive treaty was replaced by more manageable but less ambitious incremental approaches. The Baruch plan was eventually followed by a series of independent control measures for nuclear arms (the Limited Test Ban Treaty, the Anti-Ballistic Missile Treaty, the Strategic Arms Limitation Talks, and now the Strategic Arms Reduction Talks). The failure of ITO left the less expansive GATT in place. And the LOS, although not subsequently replaced by smaller treaties, has nonetheless established a patchwork of formal and informal norms regarding exclusive economic zones, fisheries and other contentious issues²².

Pursuit of a Law of the Atmosphere would probably result in a long delay until agreement is reached; even after signature, the agreement could easily collapse owing to defection. But although LOA or

other comprehensive approaches may not be flexible over time, such strategies deserve more attention than they have received because they incorporate a wide range of issues better than piecemeal strategies²³.

A third set of proposals centres on a greenhouse tax. Most of the work in this area has been in the context of national implementation of a carbon tax as part of a broader, unspecified global control system^{23,24}. Under such a scheme, emissions of carbon dioxide (and perhaps other greenhouse gases) would be taxed to

Few aspects of climate change are well suited for strong policy actions such as bans . . . the consensus is that there are probably no acceptable near-term solutions to climate change.

make the social costs of polluting the commons internalized in the emitters' decision-making process.

The main advantages of the carbon tax are that it can be adjusted as needed to attain certain levels of abatement and it provides a mechanism for raising revenue²⁵. Coupled with agreements regarding the use of the revenue, the carbon tax might support a number of international objectives such as scientific research, inequity reduction, or investment in more efficient technologies.

But the carbon tax is significantly less flexible over time than initially appears. Carbon taxes generate large amounts of revenue: a carbon tax of \$30 per ton applied today would generate \$160 billion in annual revenue, which would then require distribution. This is five times the entire international transfer of Official Development Assistance²⁵ and nearly a third the value of world oil production²⁶. A higher tax would yield even greater revenue (although the revenue of course depends on the quantity of emissions as well as on the tax level).

The problem is that nations use dramatically different tax and spending policies within which a carbon tax — or any other large source of revenue — would have to be accommodated. This global warming strategy would fix progress to the rate of tax reform in the slowest nation, because it is unlikely that a nation will simultaneously use a significant carbon tax and retain its old tax policy. (Incidentally, many governments will be nervous about such large sums of money raised in the context of slowing global warming because that money will be a highly visible target for international transfer.)

A fourth approach is a global system of tradeable permits^{15,27,28}. Under a variety of possible allocation schemes (see, for example, ref. 15) emitters would be issued

permits which they could then trade on open markets. The system works by creating a property right (the pollution permit), whose value yields economic incentives to reduce emissions because lowered emissions create a windfall of permits which can then be sold. (Or, reduced emissions would require the purchase of fewer permits.) By controlling the number of permits, environmental goals are guaranteed to be met (as long as the systems is well monitored and enforced).

Experience with tradeable permits in domestic environmental policy is limited and the lessons are mixed but promising^{29,30}. In contrast to command-and-control strategies, tradeable permits or other market-based approaches may achieve environmental protection at lower cost^{31,32}. The authors of a recent study²⁸ persuasively argued that the tradeable permits approach is the most economically efficient of several strategies to slow global warming. Further, the sale of excess permits by developing countries may provide an important source of economic development assistance.

Applied internationally, the tradeable permits approach has two problems³³. First, it will be politically difficult to convince many nations to create pollution rights. Many find pollution rights morally indefensible, and developing countries will equate emissions rights with development rights because emission of greenhouse gases is typically considered a part of normal economic development. Marketing those rights will allow the fortunate to 'buy up the environment'. But some of these equity-based fears may be addressed if the permits are leased³⁵, thus reducing the chances that permits will be hoarded by the rich.

The system's second problem also stems from its focus on property. Because property rights are inherently revalued and redistributed whenever the terms of the agreement are substantially changed, property owners will form another coalition which could selectively block changes in the level of abatement (that is, the number of permits) and the distribution of burden-sharing (that is, allocation of the permits). The tradeable permits system must therefore be considered inflexible over time.

In theory, both of the market mechanisms considered here — greenhouse taxes and tradeable permits — allow national styles of governance to prevail. Governments could pay the taxes or trade the permits and then use their choice of domestic strategies.

A final approach is a climate convention-protocol system. Advocates of this approach support the development of a framework climate convention to outline the principles and objectives of slowing global warming. A subsequent protocol to the convention would set numerical

targets for the control of greenhouse gas emissions. Protocols would also specify particular issue linkages such as funding mechanisms and technology transfer¹⁷. As noted earlier, the convention-protocol system could be hybrid in nature; for example, Nitze¹⁴ recently proposed a convention-protocol structure with quasi-command-and-control components (such as a suggested requirement that OECD nations increase energy efficiency by a certain amount each year).

The convention-protocol perspective focuses on achieving numerical targets which are codified in the language of the protocol. Uniform targets could be established for all countries or may vary according to metrics such as a country's per capita income, GNP, population or area. Variability of control levels allows some equity concerns to be addressed. In the case of acid rain, for example, the sulphur-dioxide protocol to the Long Range Transboundary Air Pollution (LRTAP) convention varies control levels according to ability to pay¹⁵. In the case of stratospheric ozone depletion, the Montreal Protocol provides a ten-year delay on control measures for developing countries.

The convention-protocol structure is probably the most politically acceptable of the five classes of agreement reviewed here. The popularity of this approach is due in part to the success of the Montreal Protocol on Substances that Deplete the Ozone Layer. Recent amendments and adjustments to the protocol to expand and strengthen controls on ozone-depleting gases are a promising indicator of the long-term flexibility of the convention-protocol approach¹⁶.

The recent strengthening of the Montreal Protocol also established a \$240 million, three-year fund to aid the shift of developing countries from ozone-depleting to ozone-benign chemicals, which serves as evidence of the convention-protocol's capacity for coupling environmental requirements to other incentives.

Weaknesses

Although the Montreal Protocol has been effective and inspiring thus far, the agreement also reveals several weaknesses of the convention-protocol approach. First, the Montreal Protocol is fundamentally an agreement on allowable emission of ozone-depleting substances and contains few provisions for linkage across issues. There are trade sanctions built into the protocol, but they are limited to ozone-depleting substances and products. The fund is a promising route to issue linkage, but it is limited only to grants and loans related to replacement of ozone-depleting substances.

Second, and perhaps most important, is that the nations' incentives for signing the Montreal Protocol were the benefits of

limiting ozone depletion (and avoiding trade sanctions) whereas the costs of switching to less ozone-destructive emissions were relatively small. An indicator of this is that barely five years passed from the signing of the framework convention in 1985 until the protocol was strengthened in 1990 completely to phase out all the egregious emissions that contribute to ozone depletion.

Global warming is quite different from ozone depletion because the needed issue linkages are both more extensive and expensive. The differences do not necessarily mean that the convention-protocol approach is unable to address adequately the global-warming issue, only that the Montreal Protocol model is incomplete. Applied to global warming, the protocol approach requires specification of the terms of agreement. Although the stringency and structure of the agreement can be changed (and have been in many precedents, including the Montreal Protocol) it remains to be seen if this will be sufficiently flexible given the large number of issues to be addressed.

In sum, the convention-protocol approach is favourably regarded, but it is unclear if the approach will allow the range of important issue linkages needed to gain widespread acceptance among nations. Notably, if the Montreal Protocol 'script' is followed then a global-warming agreement will consist of emissions quotas for nations and, perhaps, a small fund to help developing countries. More extensive incentives will probably be required to gain greater leverage over global warming. If those incentives are created by linking other issues to a climate convention-protocol structure then the issue linkages would have to be written into the structure of the protocols which will constrain long-term flexibility.

An alternative model for a climate agreement is to use the model of GATT, the postwar agreement which has progressively lowered trade barriers and expanded world trade over the last 40 years. GATT's core has been the liberal principle that free trade is beneficial to all nations; reciprocal concessions of economic national protectionist policies would (and has) yielded to positive sum benefits of freer trade.

The agreement itself is a set of rules governing acceptable trade-related practices; over time the application of those rules has changed through a semicontinuous process of negotiation 'rounds'. Each round has begun with abstract goals, which have solidified during the negotiation process. Some rounds resulted in general principles or agreements, whereas others yielded specific product-by-product reductions in tariffs for tens of thousands of tradeable goods. Overall, tariffs have reduced by three-fourths in the postwar period.

Applied to the climate case, a General Agreement on Climate Change (GACC) would be organized around the core principle of managing the rate and extent of global climate change. Core agreements would consist of allowable contributions to global warming over specific periods of time¹⁷, which might be set for individual nations or groups of nations. In some cases, guidelines on greenhouse-related activities would be more appropriate than rigid quotas. Other agreements could include income transfers, so-called 'debt-for-nature swaps', or other arrangements as required to address the concerns of global warming and to ensure wide involvement in the international effort to slow such warming.

Negotiations

Unlike the protocol approach — where the type of issue linkage (for example, a particular financial mechanism) would be written into the structure of the agreement — the GACC structure would not specify what types of agreements and linkages must be included in each round of negotiations. The shell of the General Agreement would contain only the procedures for codifying and changing the terms of the agreement as well as the general objective of slowing or even reversing climate change. Negotiators would determine the issue linkages relevant to each round; presumably these would change as the economic, technological, scientific and political contexts change. Notably, the issues will change as enticements needed to gain widespread agreement change.

At this writing, the GATT 'Uruguay Round' — which has addressed a wide range of trade-related issues such as intellectual property rights, the service sector, agricultural subsidies and other non-border impediments to free trade — is likely to end with failure. Why choose the GATT model when GATT and world trade seem in disarray? First, of interest is the model of iterative, highly flexible negotiations, not GATT itself. If the argument that climate change is fundamentally different from other international environmental issues is persuasive, such models which allow flexibility for negotiation should have a premium.

Second, GATT is in disarray because its core principle — the liberal theory of barrier-free trade — is not universally accepted¹⁸. In times of economic boom and US dominance over the world economy, there were clear incentives to reduce some trade impediments such as border tariffs. GATT prospered under those conditions. Yet nations have always sought measures to serve national interests at the expense of international goals such as free trade. The present GATT impasse on the issue of domestic agricultural subsidies precisely reflects the tension between national and inter-

national goals, with the former clearly emerging as dominant.

In the case of GACC, a challenge will be the enhanced accommodation of domestic constraints on international cooperation. The core GACC principle is the need to slow climate change through agreed national targets. GACC will provide maximum flexibility in setting those targets and relevant issue linkages, but there is little that any agreement can do if nations do not agree with the premise that slowing climate change is worth sacrifices elsewhere.

Although the GATT model is useful, its applicability to the proposed GACC structure is not complete. The challenges of global warming are quite different from the goal of reducing trade barriers. Notably, GACC will have to surpass GATT in its accommodation of the interests of developing countries. The main reason for not specifying the particular types of agreement that should form a GACC round is that diverse interests — especially those of developing countries — will change over time. For example, the debt crisis — which has spawned the interest in 'debt for nature swaps' — was relatively nonexistent 15 years ago. The issues that will emerge as important links to environmental protection and the participation of developing countries in the future are very unpredictable.

GACC may be best suited to address issues which are politically difficult (or impossible) for some countries. GACC could enable smaller bilateral or multilateral arrangements which would be sanctioned by the agreement but which would not require the participation (or ratification) of all signatories. For example, the debate over the long-term benefits of high population density notwithstanding³⁹, many nations favour incentives to slow population growth in developing countries as part of a global strategy to address the root causes of greenhouse emissions. But, in the case of the United States, inclusion of such controls in a GACC round could block ratification by Senate members opposed to population control. Smaller agreements, therefore, could refer to only those nations for which the political objections are less serious.

These smaller negotiation groups could also help congregate actors of common interest and help smooth the complicated process of negotiating agreement among so many actors of diverse interests^{40, 41}. In general, simulated climate negotiations support the organizing principles of GACC, highlighting the need for: (1) continuous negotiations; (2) coalition formation; (3) evolution of the agreement over time; and (4) enticements (side payments) for countries which would not otherwise sign the agreement (E.A. Parson and L. Susskind, personal communication).

How might we get to GACC? Between

Greenhouse emissions — how long can we wait for policy action on global warming?

now and UNCED, it is essential that all possible participants in such an agreement negotiate the rules by which the agreement would be created and modified. These negotiations would also determine which changes to the agreement would require re-ratification and which would not. As with GATT, the GACC structure would also include processes for dispute resolution. Similarly, these negotiations would have to establish a framework for enforcement. Reciprocity ("If you obey, I will obey") is readily available, but it is weak (indeed reliance upon poorly enforced reciprocity remains a major problem for GATT). Trade sanctions or other penalties have been proposed as an enforcement measure (and were included in the Montreal Protocol), but nations will fear the loss of sovereignty and manipulation of trade. Enticements to comply (such as issue linkages described above) may offer the best mechanism for enforcement. The negotiations over the GACC structure would have to address enforcement provisions (although partial enforcement procedures might not be established and, instead, left to each round of negotiations and/or the dispute resolution process).

For any greenhouse agreement, even the process of determining its structure will encounter difficulties due to large numbers of participants and diverse interests. But it seems likely that the GACC structures could be ready by 1992 and could be adopted in conjunction with UNCED in an effort to underscore the importance of issue linkages across environment and development.

In addition to rules for changing the agreement, the process of creating GACC must also include support activities for

future GACC rounds and agreements. Varied scientific activities provide essential input into any climate agreement and many coordinated international research programmes already exist in this area, including those conducted under the aegis of the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO).

By far, the largest amount of climate-related science is still directed by national programmes, and this pattern will presumably continue in the future. But some activities will have to be replaced or duplicated at the international level as part of a greenhouse regime, especially the monitoring of emissions and greenhouse-related activities. GACC could serve as the logical institutional setting for a monitoring programme, although an appropriate UN agency — probably the WMO — would presumably oversee the scientific activities. Actual monitoring could continue as a predominantly national or multilateral activity, but international oversight would help ensure accuracy (and compliance). Also, international funding and orchestration could be provided for monitoring activities not adequately covered by national programmes, to help prevent underinvestment in the needed greenhouse accounting system.

Monitoring activities could fill part of the need — identified by Thacher⁴² and others — for coordinated scientific and policy programmes. Perhaps scientific programmes under GACC could also oversee experiments with 'technical fixes' because solving certain aspects of climate change are highly relevant to the terms of the agreement.

Monitoring activities may prove among

the most important contributions of GACC (or any other greenhouse agreement). Students of international institutions argue that transparency of national activities contributes to institutional effectiveness^{16, 43, 44}. Informal norms arise which constrain behaviour when the activities of nations are visible to all. In the case of global warming, transparency would entail: (1) an accounting of each country's emissions of each greenhouse gas; (2) the steps taken to control those emissions; and (3) the efficacy of those steps.

The core of the accounting system would be the monitoring activities described above; in short, a GACC-sponsored monitoring programme would play an essential role in building transparency into the GACC regime. Monitoring under GACC could also make important contributions to the science of global change. Many monitoring activities (for example, the rate and degree of deforestation) require intrusive measures. Nations otherwise reluctant to submit themselves to such measures might be more willing if coupled to the incentives afforded by participating in GACC.

Even if GACC is not pursued, a similar monitoring programme should be explored. At present there are extreme problems in monitoring and verifying emissions of some sources of greenhouse gases (especially for methane, a strong greenhouse gas)^{45, 46}; sanctioned monitoring activities may reduce these problems and increase the prospects for eventually including those uncertain sources in a greenhouse agreement.

Conclusions

Although there are many tradeoffs, clearly an international agreement to slow global warming must gain widespread acceptance if it is to command substantial leverage over the greenhouse effect. Further, such a regime must be adaptive because: (1) cooperation to control global warming cuts across many issues; (2) those issues will change over time (and have in the past); (3) the scientific assessment of global warming will change over time; and (4) the political interests — including the perception of the risks of global warming — will change. To different degrees, the various structures under consideration are able to meet the criteria of political acceptability and flexibility over time. An alternative which better meets these criteria is a General Agreement on Climate Change (GACC), patterned on the iterative, flexible negotiation structure of the General Agreement on Tariffs and Trade (GATT).

By far, the most popular structure under consideration is the convention-protocol approach. This is amply evident in the work of the Intergovernmental Panel on Climate Change¹¹. As a framework, there are no inherent problems with

the convention-protocol approach. Indeed, the GACC structure proposed here might be institutionalized through a climate convention.

But the main difference between GACC and the convention-protocol approach that many envision is that GACC does not specify the types of issue linkages in its structure. It is conceivable that each round of GACC negotiations would involve different issues, as needed to entice participation of certain states and to address the root causes of greenhouse emissions. In practice, some issues would remain the same from round to round — such as financial and technology transfer — whereas others would appear, change and disappear from the negotiations as necessary to bring nations to the bargaining table and reach agreement.

As such, GACC is similar to most existing environmental agreements because the terms of the agreement would change over time. The departure from the convention-protocol approach is made because, unlike virtually all existing environmental agreements, the issues related to reaching agreement on global warming are widespread and highly variable over time.

Following Sebenius⁴⁰, the GACC approach is in-between the issue-by-issue structure of the convention-protocol approach and the comprehensive struc-

ture of a Law of the Atmosphere. GACC would eventually allow a comprehensive approach to climate control, but it would begin modestly and evolve over time. As with many visions of a convention-protocol approach, GACC would allow nations to pursue their own domestic policies — command-and-control, market-based strategies, or hybrids — provide that the terms of the agreement are met.

There is still cause for considerable pessimism regarding the prospects for a climate agreement. The many diverse interests will pose an extreme challenge for international cooperation. But the GACC structure — because it is the least specific — can accommodate diversity and may be the best approach for creating an adaptive greenhouse regime that commands widespread acceptance among nations. Still, even this approach will be extremely difficult to negotiate and ratify.

Even if GACC is initiated, we should expect failures and should structure GACC as such that the entire process does not collapse if one or more rounds do not yield agreement. Confronting this issue of failure and adaptability is one of the largest challenges of global warming. □

David G. Victor is at the Department of Political Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

- Intergovernmental Panel on Climate Change. *Climate Change: The IPCC Assessment* (Cambridge University Press, 1990).
- Skolnikoff, E. B. *Foreign Policy* **79**, 77–93 (1990).
- Hardin, G. *Science* **162**, 1243–1348 (1968).
- Stone, D. *Policy Paradox and Political Reason* (Scott and Foresman, 1988).
- Olson, M. *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard University Press, 1965).
- Oye, K. in *Cooperation Under Anarchy* (ed. Oye, K.), 1–24 (Princeton University Press, 1986).
- Marland, G. et al. *Estimates of CO₂ Emissions from Fossil Fuel Burning and Cement Manufacturing* (Oak Ridge National Laboratory, 1989).
- Ministerial Declaration of the Second World Climate Conference (7 November 1990), par. 12.
- United Nations General Assembly Res. 43/196 (1988).
- Subak, S. *Research Memorandum* (Stockholm Environment Institute-Boston, 1990).
- Intergovernmental Panel on Climate Change. *Climate Change: The IPCC Response Strategies Report* (Island Press, in the press).
- Smil, V. *Pop. Dev. Rev.* **16**, 1–29 (1990).
- Houghton, R. A. et al. *Tellus* **39B**, 122–139 (1987).
- Young, O. R. *Int. Organ.* **43**, 349–375 (1989).
- Grubb, M. *Greenhouse Effect: Negotiating Targets* (Royal Institute of International Affairs, 1989).
- Young, O. R. in Rosenau, J. N. (ed.) *Governance without Government: Change and Order in World Politics* (Cambridge University Press, in the press).
- Haas, E. in Krasner, S. D. (ed.) *International Regimes* (Cornell University Press, 1984).
- Ausubel, J. H., Grubler, A. & Nakicenovic, N. *Climate Change* **12**, 245–263 (1988).
- Ausubel, J. H. *Nature* (in the press).
- Goldemberg, J. et al. *Energy for a Sustainable World* (Wiley-Eastern, 1988).
- Zaelke, D. & Cameron, J. *Am. Univ. J. Intern. Law Policy* **5**, 249–290 (1990).
- Sebenius, J. K. *Discussion paper* (Global Environmental Policy Project, John F. Kennedy School of Government, Harvard University, G-90-10).
- Gaskins, D. & Stram, B. *Discussion Paper* (Global Environmental Policy Project, John F. Kennedy School of Government, Harvard University, in the press).
- Manne, A. S. & Richels, R. G. *Energy J.* **11**, 51–74 (1990).
- World Bank. *World Development Report 1990* (Oxford University Press, 1990).
- Assuming world oil production is ~60 million barrels per day (mbd), at \$20 per barrel. British Petroleum reported 1989 production at 63mbd and spot crude (West Texas Intermediate) at \$19.68. British Petroleum. *BP Statistical Review of World Energy* (BP, 1990).
- Stewart, R. B. & Wiener, J. B. *Am. Enterprise* **1**, 75–80 (1990).
- Epstein, J. M. & Gupta, R. *Controlling the Greenhouse Effect: Five Global Regimes Compared* (Brookings Occasional Papers, 1990).
- U.S. Congressional Research Service, 1989. *Using Incentives for Environmental Protection: An Overview* (B9–360 ENR, 1989).
- Opschoor, J. B. & Vos, H. B. *Economic Instruments for Environmental Protection* (OECD, 1989).
- Tietenberg, T. H. *Emissions Trading: An Exercise in Policy Reform* (Resources for the Future, 1985).
- Project 88: Harnessing Market Forces to Protect our Environment (US Senate, Offices of T. E. Wirth and J. Heinz, 1989).
- Victor, D. G. *Discussion Paper* (Global Environmental Policy Project, John F. Kennedy School of Government, Harvard University, G-90-07).
- Nitze, W. A. *The Greenhouse Effect: Formulating a Convention* (Royal Institute of International Affairs, 1990).
- Fraenkel, A. A. *Harvard Int. Law J.* **30**, 447–476 (1989).
- Report of the Second Meeting of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer* (UNEP/OzL Pro.2/3, 1990).
- Eckaus, R. S. *Working Paper* (MIT Center for Energy Policy Research, 90-022, 1990).
- Gilpin, R. *The Political Economy of International Relations* (Princeton University Press, 1987).
- Simon, J. L. *Pop. Dev. Rev.* **15**, 323–332 (1989).
- Sebenius, J. *Int. Security* (in the press).
- The Salzburg Initiative* (International Environmental Negotiation Network, Harvard Law School, 1990).
- Thacher, P. S. *Colorado J. Intern. Envir. Law Policy* **1**, 101–126 (1990).
- Chayes, A. & Chayes, A. H. in Mathews, J. T. (ed.) *Preserving the Global Environment: The Challenge of Shared Leadership*, 280–308 (Norton, New York, 1990).
- Keohane, R. in Krasner, S. D. (ed.) *International Regimes*, 141–171 (Cornell University Press, 1984).
- Victor, D. G. *Policy Sci.* (in the press).
- Fischer, W., di Primio, J. C. & Stein, G. A. *Convention on Greenhouse Gases: Towards the Design of a Verification System* (Forschungszentrum Julich GmbH, Programm-gruppe Technologiefolgenforschung, 1990).

Non-profit profits from animals

A Californian zoo much admired for the manner in which animals are displayed may have hit on an even more important secret — that of running a scientific institution without begging for public funds.

San Diego

THIS city is almost as well-known for its zoo as for anything else. By the 1950s, the San Diego zoo had won a generally envied reputation for the style in which its animals were exhibited. Relics of that period survive. But the zoo is now in the throes of remodelling itself so that its animals will be seen in simulations of their natural habitats, while the collection itself is being trimmed to accord with newer conservation doctrines. Even so, San Diego's chief claim on the attention of other zoos may be that it does all this without substantial subvention from public funds.

San Diego is more than a zoo-city, of course. It is the principal West Coast naval base of the United States (which is now mourning the first US land casualties of the Gulf War among marines from their base near the city). The city is also proud of the ninety or so biotechnology companies founded since Hybritech pitched its tent here in 1978. Then, a few miles further north, there is a clutch of distinguished research institutes which now almost dwarf the first of them, the Scripps Institution of Oceanography. (Chronologically, the Salk Institute came next, followed by the Scripps Clinic and Research Foundation.) The San Diego campus of the University of California is also working its way up the league table of federal grant recipients — last year it had reached sixth place. But the zoo is even more widely known, or at least more widely appreciated.

The zoo and its associated wildlife park is an astonishing enterprise, with an annual budget of more than \$70 million. Stung by San Diego's reputation, other city-centre zoos tend to remark on the size and splendour of its 100-acre site, a mesa (where the car park is) scored by small canyons now being increasingly used as the sites of animal exhibits. But the reality is even better than it seems on paper. Originally on the outskirts of a small city, the zoo has seen one of the most rapidly growing cities in North America flow around it until it is no further from the city centre than are the principal art museums elsewhere.

San Diego's success as a zoological institution rests squarely on its willingness to be a part of the robust entrepreneurial traditions of North America. The zoo's hardcore support comes from the 180,000 families who have taken out membership of the Zoological Society of San Diego, which allows them open access (and privi-

leged but paid access to special events). At \$50 a year, the zoo membership may be "the best buy in San Diego". There are also 80,000 youth members.

But out-of-town visitors are the hard core of the business. There are close on 2.5 million a year at the zoo, and more than half as many again at the wildlife park. As these things go, a day at the zoo is cheap at \$10.75, but San Diego reckons that what visitors buy (food and souvenirs) is even more valuable. And "the advertising is brilliant", one supporter says. Like any other business, the zoo changes advertising agents when that seems prudent.

The only regular contribution to costs from public funds is from a surcharge on property taxes levied by the city, which raises \$2.5 million a year, ostensibly for horticulture at the zoo (which can mean landscaping). San Diego is also always on the lookout for other sources of funds — researchers apply for grants like other people, notably for work at the Center for Research on Endangered Species (which is also building up a substantial endowment fund). Just now, the zoo is embroiled in a row about its allegedly inattentive use of funds granted by the state of California for the retraining of zoo keepers.

The zoo's now-careful planning seems a far cry from its impulsive creation, in 1916, by the migrant east-coast physician Dr Harry Wegeforth, who is said to have found some caged lions left over from an international trade exhibition, at which they had been used to entertain the crowds, and to have persuaded a local benefactress (name of Scripps) to give them a more fitting home in a dark concrete cave with natural barriers, rather than bars, to keep the people away from them.

Geography being what it is, the zoo has sought out from that start exotic species from South America and the eastern seaboard of Eurasia (no further away). Its tradition of providing natural settings for its animals (there are 3,500 at present) stems from the beginning, and is as much a sign that the zoo grew empirically as a measure of the generosity of its site. Its distinctiveness is the work of a handful of Wegeforth's successors, of whom the latest is David Rice, an architect turned zoo-designer, with a flair for turning the sides of mini-canyons into replicas of animals' habitats in miniature.

Thus San Diego already boasts of what it calls Tiger River, a steep track lined with

tropical vegetation to simulate a rain forest (the canopy is not apparent) at the bottom of which is an enclosure for half a dozen tigers (the rare Siberian tiger lives elsewhere) which is perhaps 50 feet across, twice as long and also 100 feet high. The tigers, with their weight, have made a nice simulation of semi-arid India of their narrow patch. Rice is also somewhat surprised that a troupe of sun bears have so quickly worn through the green turf with which his enclosure had provided them.

Rice is now at work on an ambitious gorilla enclosure, intended as a slice of hillside that will be covered with appropriate vegetation. The site is now covered with a small army of people bent double with planting tools and with cranes intended to lift mature trees, with their roots encased in huge neat wooden boxes, into the holes that have been dug for them. Rice is about the only one on the site who is unperturbed that, after eighteen months, the whole enclosure is due to be occupied on 23 March.

These developments cannot but enhance San Diego's reputation as a show-place, but the zoo is also being nudged in other directions by the climate of the times. Taking animals from captivity, except when that is the best means of ensuring the survival of their species, is naturally frowned upon, which is why the zoo has concentrated so much attention on breeding in captivity. The general principle is that each species should be represented by enough members to give them a reasonable chance of maintaining themselves. There have been notable successes — San Diego is proud of the southeast Asian clouded leopards (*Neofelis nebulosa*) it has bred, as well as of its black rhinos. There are also hopes that the time is fast approaching when it will be possible to begin releasing Californian condors bred in captivity to the wild, and even that the Andean condors which have grown up in the zoo will also be released.

Yet it has in common with the meanest zoos the trick of provoking the anthropocentric question why natural selection generated such a rich variety of magnificent animals that are nevertheless endangered. And it is doubtful whether even the splendid new exhibition areas planned will rid all visitors of the embarrassment that comes from looking at animals placed there for no other purpose than to be looked at.

John Maddox

Climate plumbs the depths

John Sass

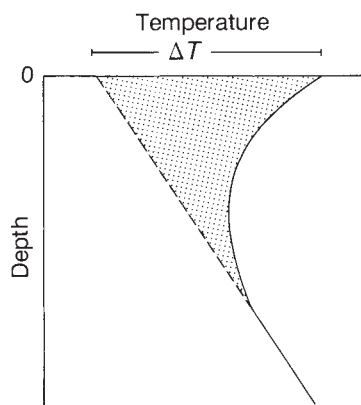
If you wanted to learn about the past climate, your last inclination might be to measure the temperature of soil deep beneath your feet. Yet in the absence of disturbances introduced by convection of groundwater, surface water or local topography, the upper few hundred metres of the Earth store a record of temperature change for the past century or two. This fact is well known to those measuring the Earth's heat flow, for whom the climate signal has been a source of annoyance. But the implications for climate research have not been fully appreciated, which is why a special session was convened at the recent meeting* of the American Geophysical Union.

In the absence of any disturbances, thermal energy from the Earth's deep interior will flow steadily to the surface across a temperature gradient that is a function of the total heat flux (a tiny fraction of the radiant heat from the Sun) and of the thermal resistance of the geological medium. By studying the temperature gradient in boreholes, geophysicists can learn about the heat sources and sinks involved in such geological processes as volcanism, faulting, subduction and basin evolution. But by perturbing the mean surface temperature, climate variations alter the near-surface heat flow. Effectively, the disturbance is propagated down into the crust, leaving a record which can now be recovered (see figure).

Meteorological observatories are few in number and are mostly located near population centres, sites of surface modification and, consequently, of anthropogenic temperature change. On the other hand, there are many thousands of boreholes worldwide from which equilibrium temperature profiles can be obtained. Where these are close to meteorological observatories, their temperature profiles can be compared with surface temperature records and used to extend records back in time or to recognize emerging trends not apparent in the noisier surface data set. Temperature profiles from locations far from observatories can significantly extend the spatial coverage of the limited observatory data base.

Until recently, research on climatic implications of ground-temperature measurements was concentrated in arctic regions, because of the stability ensured by the thick, impermeable permafrost layer and because the temperature disturbance was larger there during the past century. The focus of the past few years on global warming and the controversy surrounding global climate models and other

measures of climate change, have prompted heat-flow researchers to re-examine thermal data from low- and mid-latitude boreholes in an effort to identify curvature in the upper parts of temperature profiles and to provide estimates of the size, nature and duration of the surface-temperature disturbances associated with the observed curvature. Heat-flow workers have found evidence of warming in the Canadian Shield, the US Great Plains, Cuba, Southern Africa, Antarctica and elsewhere. In some areas,



Schematic representation of the effect of surface warming ΔT on the temperature profile from a borehole (solid line, perturbed profile; broken line, unperturbed). The perturbation typically penetrates 30–40 metres after a decade, 100–150 metres after a century.

curvature is absent from the temperature profile, or, if present, can be explained in terms of microclimates resulting from nearby lakes or changes in land use. In other areas, there is evidence of cooling, as expected from global climate models.

Although the ground filters out high-frequency variations in atmospheric temperature, ground and air temperatures should follow one another closely over periods of months to decades. Most observatories monitor only air temperatures, so that the complicated heat-exchange process between air and ground has not been well-characterized. Where the two temperatures have been measured, mean annual ground temperatures are usually higher than the air temperatures — by a degree or so in temperate zones, and by several degrees in areas of deep seasonal snow cover.

Studies of global air-temperature variations over the past century indicate that the climate has warmed by about 0.5 °C over that period. This disturbance should be readily apparent in the upper 200 m of the temperature profile of the Earth and, indeed, many contributors to the special session presented data supporting warming of this magnitude or greater. In general, the temperature history inferred from

the borehole measurements in a given region concur with the averaged records of air temperatures from the same region.

Data from Alaskan arctic regions illustrate the great sensitivity of the permafrost layer to surface-temperature perturbations that have occurred over the past century (A. H. Lachenbruch, US Geol. Surv.). The total temperature rise of the Alaskan north slope during the past century is between 2 and 4 °C. Similar increases have been inferred from borehole temperatures in the Canadian Arctic and the Canadian Shield (A. M. Jessop, A. S. Judge, T.J. Lewis, Geol. Surv. Canada; J.-C. Mareschal, H. Beltrami, Univ. Québec). Temperature rises in the range 0.5–2 °C during the past century were reported from the north-central United States (W. D. Gosnold, M. Bauer, Univ. North Dakota), The Basin and Range Province of Utah (T. J. Chisolm, D. S. Chapman, Univ. Utah), and Southern Africa (H. Pollack, Univ. Michigan).

Although climatic effects can be recognized in temperature profiles (A. E. Beck, Univ. Western Ontario), the inversion of temperature–depth data to extract a measure of surface temperature as a function of time will not yield an unambiguous thermal history without additional constraints. It is possible, however, to distinguish between, say, a temperature step and a linear increase over time in surface temperature by a careful study of the curvature in the temperature profile. Formal mathematical inversion procedures can be used to establish limits for the resolution of the method as a function of the accuracy of temperature and depth measurements (G. D. Clow, US Geol. Surv.; D. R. MacAyeal, Univ. Chicago).

A cautionary note was sounded (D. D. Blackwell, Southern Methodist Univ.) in that the locations of many boreholes are biased towards sites where the curvature characteristic of surface warming is caused by other factors (as on south-facing slopes and close to bodies of water).

The special session on climate change and borehole temperature provided a lively forum, both for comparing notes among heat-flow specialists, and for dialogue with meteorologists and climatologists. From the data presented, it was clear that surface temperature increases of up to a few degrees during this century are widespread, but not ubiquitous. Theoretical studies provided useful input on the sensitivity and limitations of extracting climate information from temperature profiles. The discussions defined a framework for a systematic study of the thousands of temperature profiles available. We can expect to see important contributions to climate research from this source soon. □

John Sass is at the US Geological Survey, 2255 North Gemini Drive, Flagstaff, Arizona 86001, USA.

*Climate Change Inferred from Borehole Temperature, AGU Fall Meeting, San Francisco, 3–7 December 1990.

Mls: makes a little sense

Charlie Janeway

THE mysterious nature of the minor lymphocyte stimulating¹⁻³ (*Mls*) genes once led an immunologist to quip that "*Mls* stands for makes little sense". That may change now that several groups, four of which describe their findings in this issue, have independently produced evidence that *Mls* and related genes specifying self-superantigens are encoded by retroviruses⁴⁻⁹.

The *Mls* genes were named for their ability to stimulate remarkable proliferative responses of CD4 T cells in mixed lymphocyte cultures between strains identical at the major histocompatibility complex (MHC), the only other known stimulus for such responses. It has become clear that *Mls* and other superantigens work by binding simultaneously to an MHC class II molecule and all T-cell receptors encoded by one or more V β gene segments (see figure), accounting for the strength of the proliferative response. Alternatively, the presence of self-superantigens in a given mouse can be detected by the absence of mature T cells bearing receptors encoded by that V β gene segment. Cells bearing these auto-reactive receptors are generated in the thymus, but they are deleted before maturation to maintain non-responsiveness or tolerance to self-superantigens. The stimulation and/or deletion of T cells expressing a particular V β is diagnostic for the presence of a self-superantigen or, as I prefer, a V β -selective element (Vbse).

The interaction of the T-cell receptor with Vbse seems to be radically different from the interaction of the same receptor with protein antigens. The specificity of antigen recognition by the T-cell receptor results from the involvement of all the variable receptor elements in the recognition of a peptide fragment of antigen bound in the cleft of a specific MHC molecule¹⁰. By contrast, the high frequency of T cells responsive to *Mls* occurs because recognition of Vbse depends only on the portion of the receptor encoded by one or more of the 20 V β gene segments (see figure). The crucial region for T-cell receptor interaction with Vbse lies on the lateral face of the V β domain¹¹⁻¹³, a region not involved in conventional antigen recognition¹¹. Vbse also require an MHC class II molecule (usually I-E) to produce their effects, although little influence of MHC polymorphism is seen in either T-cell activation or clonal elimination by Vbse. As in other responses to class II MHC, the T cells that respond to *Mls* bear CD4, and depletion of CD4 T cells generally eliminates such responses¹⁴.

Earlier work pointing to a genetic linkage of a Vbse with an endogenous retro-

viral genome⁴ is strongly supported by the new studies, which make a convincing case that Vbse are encoded by retroviral genes. This case consists of four parts. Marrack *et al.* (page 524 of this issue⁵) describe a novel Vbse that is vertically transmissible in the milk of mice known to pass active mouse mammary tumour virus (MMTV) from mother to offspring by this route. Mice cured of MMTV infection by foster nursing do not show clonal deletion, implying that the Vbse is encoded by this virus. This is the only Vbse transmitted in this way; no direct conclusions about genes encoding endogenous Vbse can be drawn from this finding.

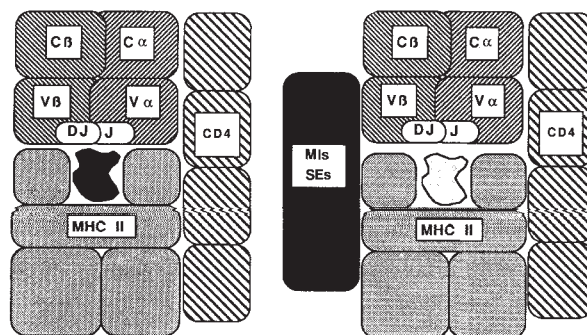
A connection between endogenous retroviral sequences and inherited Vbse is made independently by three other groups, whose reports appear on pages 526, 529 and 531 of this issue⁶⁻⁸. A number of Vbse detected by intrathymic deletion and/or stimulation of cloned T-cell lines are genetically inseparable from endogenous retroviral genomes, all of which are members of the mammary tumour virus (*Mtv*) family (see table). Moreover, cell lines which stimulate Vbse-responsive T cells express the appropriate *Mtv* transcript, whereas genetically identical cell lines which fail to stimulate such T cells lack expression of this transcript⁷. Finally, a B-cell tumour containing a member of a different retroviral family, a defective murine leukaemia virus (MuLV) that induces an AIDS-like disease in mice, also expresses a Vbse⁹. Only B-lymphoma cells that express the viral *gag* p30 fusion protein on the cell surface express the Vbse, and, more importantly, antibody to p30 inhibits stimulation by this B lymphoma.

This last finding directly implicates a retroviral protein in the T-cell response to Vbse, and definitive proof that retroviral genes determine Vbse structure will apparently follow soon through cloning of the relevant gene and the demonstration that expression of this gene leads to Vbse biological activity¹⁶. Preliminary evidence suggests that 3' sequences, most probably the open reading frame in the *Mtv* long terminal repeat, encode *Mls*-I and the mouse mammary tumour virus Vbse⁵. As reported at the recent UCLA symposium at Keystone, Colorado, Acha-

Orbea, MacDonald and colleagues have shown that mice transgenic for *Mtv*-2 also have a new I-E-dependent Vbse. These data further support the case for retroviral coding of Vbse.

Assuming that transfection experiments will confirm retroviral coding of Vbse, this finding helps to explain some of the curious aspects of Vbse biology. Biological studies of *Mls* genes had suggested that there were stimulatory and null alleles³, which correlate respectively with the presence and absence of a given retroviral genome at an *Mls* locus⁵. Thus, a null allele at *Mls* may be just that. Furthermore, some of the genetic complexity of *Mls* may be explained by the finding that similar effects on T-cell development can be imposed by different *Mtv* integrants^{6,8} (see table). Although *Mtv* genes had been thought to be expressed primarily in mammary tissue, it is clear that these genes are readily inducible in B cells by activation with lipopolysaccharide and other B-cell activators¹⁵. Interestingly, most Vbse are expressed selectively in B cells¹⁶, and their expression can be amplified by B-cell activation¹⁷. Responses of T cells to *Mls* are notoriously variable, in keeping with the inducible nature of these genes.

The long-awaited identification of Vbse as retroviral products comes as a surprise. The mouse genome may contain over 1,000 endogenous retroviral sequences. The *Mtv* integrants appear to have been part of the mouse genome for up to a million years, to be highly variable between strains, but to be stable genetic markers in breeding studies¹⁸. If Vbse are unique to the mouse, then their general biological significance must be ques-



Left: T cells respond to conventional protein antigens by recognizing a complex ligand comprising a peptide fragment of the antigen (black) bound in the cleft of a particular MHC class II molecule. All of the variable elements of the T-cell receptor are involved, especially the most variable part in the junctional sequences encoded by D and J genetic elements, as well as insertional (N) nucleotides (not shown)¹⁰. Right: by contrast, responses to *Mls* and bacterial toxic mitogens such as staphylococcal enterotoxins (SEs) involve only that part of the T-cell receptor encoded by the V β gene segment, and virtually any MHC class II molecule. Although all MHC class II molecules are believed to exist on the cell surface with bound peptides, it seems that any peptide will do for responses to SEs, so the peptide is shown in grey. Distances between structures reflect the relative importance of contacts in generating responses to the different ligands. CD4 serves as a critical coreceptor in both types of response^{14,21}.

Speedy genes

THE technique of transferring foreign genes into living cells by coating inert gold particles with DNA and literally bombarding them through the cell membrane has been fairly well documented for plants. But the successful application to mammalian somatic cells now reported by N.-S. Yang *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **87**, 9568–9572; 1990) suggests that the method may prove of value for gene therapy. Five different reporter genes were coated onto gold beads and, by accelerating them at high voltage, were successfully introduced and expressed in rodent liver, skin and muscle tissue. The method works equally well *in vitro* with a variety of cultured cell lines, and offers a potentially useful alternative to viral gene transfer.

Getting tough

RESEARCHERS at Los Alamos National Laboratory have discovered a new way of toughening the surfaces of ceramics—a vital step if these hard but brittle materials are to be put to widespread use in mechanical engineering. Silicon carbide, one of the hardest materials known, is liable to fracture owing to defects on its surface. But T. R. Hervis, J.-P. Hirvonen and M. Nastasi (*J. Mater. Sci.* **6**, 146–151; 1991) find that incorporating titanium into the SiC surface, using laser heating, has dramatic effects on the material's properties: surface cracking and flaking can be prevented almost entirely. Similar results have been achieved using ion implantation of titanium, but the new process, involving the evaporation of a thin layer of titanium onto the ceramic's surface followed by rapid heating in air by an excimer laser, is more convenient, faster and produces a surface that is in thermal equilibrium. A puzzling variation of friction with humidity apparently has no effect on the durability.

How cats purr

THE gentle purring of contented cats results from laryngeal modulation of respiratory flow, as shown by D. E. Frazer Sissom *et al.* (*J. Zool., Lond.* **223**, 67–78; 1991). The authors' comprehensive monitoring of obliging domestic cats (*Felis silvestris* f. *catus*) with oscilloscopes and sound recording equipment revealed that the sound emanates from the oscillation of the vocal folds as the cat inhales and exhales. This is independent of phonation, the pitch of which is controlled by stretching of the vocal chords, which explains why cats can purr and mew at the same time. Purr frequency is independent of age, size, weight and sex, and ultimate control is exercised by an oscillator in the brain. Parallel studies of purring in cheetahs (*Acinonyx jubatus*) and pumas (*Puma concolor*) were understandably less thorough, but yielded similar results.

Vβ-selective elements and their linkage to retroviruses

Vbse	Vβ	MHC	Chromosome	Retrovirus	Reference
Mls-1a	6, 8, 1, 9*	II, E>A	1	Mtv-7	6
Mls-2a	3	E>>A	4	Mtv-13	6
Mls-3a	3	E>>A	16	Mtv-6	6
Mls-2-like	3	?	7	Mtv-1	6, R. Abe (pers. comm.)
Etc-1 (Dvb 11-2)	5, 1, 5, 2, 11	E	12	Mtv-9	4, 7, 8
Dvb11-1	11	E	6	Mtv-8	8
Dvb11-3	11	E	14	Mtv-11	8
Unnamed	14	E	Extrinsic	MMTV	5
MAIDS B cell	5	II	Extrinsic	MuLV	9, H. C. Morse (pers. comm.)
Unnamed	17a	E	Not mapped†	?	29, 30
Unnamed	12	E	Not mapped	?	31, R. J. Hodes (pers. comm.)
Unnamed	16	E	Not mapped	?	31, R. J. Hodes (pers. comm.)
Unnamed	19a	E	Not mapped	?	R. J. Hodes (pers. comm.)

* Mls-1a is often stated to delete Vβ7 as well, but this is not always confirmed (R. Abe, personal communication).

† Non-polymorphic

tioned. One issue raised by the new findings is why retroviruses encode Vbse, and why mice retain such large numbers of these sequences in their genome. Stimulation *in vivo* with Vbse^{19,20} is known to induce widespread immune suppression, so one possibility is that retroviral Vbse facilitate viral invasion of the host by suppressing the immune response; alternatively, activation of T cells by retroviral Vbse may make them susceptible to retroviral infection, facilitating spread of the virus. The mouse may counter these strategies by retaining viral sequences which eliminate those T cells able to respond to viral Vbse.

The previously undefined nature of Mls genes has led several groups to study a similar T-cell response driven by the bacterial toxic mitogens^{14,21}. These substances cause food poisoning, toxic shock syndrome and other human disorders. They are potent mitogens for T cells whose receptors are encoded by one or a few Vβ gene segments, require class II MHC molecules to stimulate, act selectively on CD4 T cells, and cause clonal deletion of developing T cells. Thus, the bacterial toxic mitogens were thought to be possible homologues of endogenous Vbse²². If so, they seem to have achieved this result by a distinct evolutionary pathway, because they show no structural similarity to endogenous retroviral genes. An alterna-

tive hypothesis is that the bacterial toxic mitogens act on the endogenous Vbse, mimicking their effects by perturbing their structure²³. If so, then all species whose T cells respond to bacterial toxic mitogens must have Vbse, most notably humans. Although less is known about endogenous retroviruses in man, possible retroviral insertions have been reported²⁴. The new findings should allow a rapid definition of Vbse genes in several species.

The fact that Vbse may be encoded by a retrovirus and act primarily on CD4 T cells raises the question of whether the most notorious retrovirus, the human immunodeficiency virus (HIV), which causes AIDS, may also encode a Vbse that is involved in pathogenesis. CD4 T-cell depletion in AIDS could occur as follows. HIV infects CD4 T cells, remaining latent until the T cell is activated. As activated human T cells express MHC class II proteins, HIV-infected CD4 T cells would express both HIV retroviral proteins and MHC class II, allowing them to interact with T cells expressing a complementary Vβ. It has recently been shown that activated human T cells presenting bacterial toxic mitogens induce *inactivation* of cells expressing the corresponding Vβ²⁵. Furthermore, *in vivo* administration of bacterial toxic mitogens²⁶ or Mls-disparate cells²⁷ leads first to activation and then to profound depletion of CD4 T cells

- Kappler, J. *et al.* *Nature* **332**, 35–40 (1988).
- MacDonald, H.R. *et al.* *Nature* **332**, 40–45 (1988).
- Festenstein, H. *Transplant. Rev.* **15**, 62–88 (1973).
- Woodland, D. *et al.* *Science* **247**, 964–967 (1990).
- Marrack, P., Kishner, E. & Kappler, J. *Nature* **349**, 524–526 (1991).
- Frankel, W.N., Rudy, C., Coffin, J.M. & Huber, B.T. *Nature* **349**, 526–528 (1991).
- Woodland, D.L., Happ, M.P., Gollob, K. J. & Palmer, E. *Nature* **349**, 529–530 (1991).
- Dyson, P.J., Knight, A.M., Fairchild, S., Simpson, E. & Tomonari, K. *Nature* **349**, 531–532 (1991).
- Hugin, A.W., Vacchio, M.S. & Morse, H.C. III (manuscript submitted).
- Davis, M.M. & Bjorkman, P.J. *Nature* **334**, 395–402 (1988).
- Pullen, A.M. *et al.* *Cell* **61**, 1365–1374 (1988).
- Choi, Y. *et al.* *Nature* **346**, 471–473 (1990).
- Cazenave, P.-A. *et al.* *Cell* **63**, 717–728 (1989).
- Janeway, C.A. Jr *et al.* *Immun. Rev.* **107**, 61–88 (1989).
- King, L.B. *et al.* *J. Immun.* **144**, 3218–3227 (1990).
- Webb, S.R. *et al.* *J. exp. Med.* **169**, 1–12 (1989).
- Ryan, J.J. *et al.* *J. Immun.* **130**, 1063–1073 (1983).
- Coffin, J.M. in *Virology* (ed. Fields, B.N.) 1437–1500 (Raven, New York, 1990).
- Jacobsson, J., Lilliehook, B. & Blomgren, J. *Scand. J. Immun.* **4**, 181–191 (1975).
- Pinto, M., Torton, M. & Birnbaum, S.C. *Transplant.* **25**, 320 (1978).
- Marrack, P. & Kappler, J. *Science* **248**, 705–711 (1990).
- Janeway, C.A. Jr *Cell* **63**, 659–661 (1990).
- Janeway, C.A. Jr, Rath, S. & Yagi, J. *Behring Inst. Mitteilungsheft* (in the press).
- Silver, J. *et al.* *Molec. cell. Biol.* **7**, 1559–1562 (1987).
- O'Hehir, R.E. & Lamb, J.E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8884–8888 (1990).
- Kawabe, Y. & Ochi, A. *Nature* **349**, 245–248 (1991).
- Webb, S.R., Morris, C. & Sprent, J. *Cell* **63**, 1249–1256 (1990).
- Coffin, J.M. *Cell* **46**, 1–4 (1986).
- Kappler, J.W. *et al.* *Cell* **49**, 273–280 (1987).
- Marrack, P. & Kappler, J. *Nature* **332**, 840–843 (1988).
- Vacchio, M.S., Ryan, J.J. & Hodes, R.J. *J. exp. Med.* **172**, 807–813 (1990).

expressing the selected V β . So HIV might cause CD4 T-cell depletion by encoding a Vbse that is expressed by activated T cells. In this case, the slow progression of CD4 T-cell depletion in AIDS would be explained by a need for cycles of mutation in the retroviral Vbse gene, allowing depletion of CD4 T cells bearing different V β s over time. The genes of HIV are known to undergo extensive mutation during the progression of infection leading to AIDS²⁸. If this model is correct, then individual patients would be expected to lose CD4 T cells in a V β -selective fashion during progression to AIDS. This idea is supported by Morse's studies of murine AIDS as well²⁹.

The discovery that Vbse may be CRETACEOUS/TERTIARY EXTINCTION

encoded by a retroviral gene opens many possibilities. First, and most importantly, it should allow a thorough study of the structure, function and expression of Vbse, not only in mice but in humans and other species. Second, it may help to explain the pathogenesis of retroviral diseases. We can hope that these findings will eventually explain the unique phenomenology of MIs antigens and of other Vbse, telling us whether these structures play an essential role in immunobiology or are an evolutionary accident. Then, maybe, MIs will make a lot of sense. □

Charlie Janeway is in the Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

Where did it happen?

Jan Smit

THE question of whether the dinosaurs and other life forms were wiped out at the end of the Cretaceous period (66 million years ago) by a cosmic impact or by a more earthbound catastrophe may finally be answered with the discovery of glass tektites from the era in Haiti. The significance of this discovery, described by Sigurdsson *et al.* on page 482 of this issue¹, is that although there is much circumstantial evidence that a cosmic impact caused the extinction (the 'victims' and the 'bullets' have been found), there is no 'smoking gun' in the guise of a crater or of products formed by an impact. The tektites, glass drops formed in impacts, yield the missing evidence.

The Haitian tektites and plausible sites for the impacts were discussed at a recent meeting*, the first occasion for an interdisciplinary debate for two years. Those who have argued that volcanoes, not impacts, were the cause of the extinctions, remained silent at this meeting. In the past, they have been able to argue that the various geological signatures indicative of the impact mechanism (iridium contents consistent with cosmochemical, not geochemical, abundances; certain types of glassy deposits; and so on) could be justified in terms of unusual volcanic activity. Even the lingering notion that the extinction was gradual, not geologically sudden, took a setback with the reports that dinosaur taxa did not become rare before the Cretaceous/Tertiary (K/T) boundary² and that ammonites, far from declining as had been supposed, actually increased in diversity before the boundary (P. Ward, University of Washington).

The evidence from shocked minerals

has suggested that the impact usually associated with the K/T boundary layer occurred on continental crust, probably in or close to North America. The Manson crater in Iowa, now filled in, has the right age, but is too small to be the site³. The hunt for the relic impact crater has now been directed further south, following the discovery by Hildebrand *et al.*⁴ of large tektite-like particles in the Beloc site in Haiti. The Caribbean region is the only area in the world where tell-tale coarse-grained sediments are known dating precisely from the K/T boundary; such sedi-



FIG. 1 Known distribution of microkrystites (mostly altered) (○) and splash-form tektites (●) on a map of the world 65 million years ago. The number of established iridium-anomaly locations (x) is now over 105 (W. Alvarez, University of California at Berkeley). Shading, ocean crust since subducted.

ments were found, for example, in tsunami beds at Brazos River in Texas. Elsewhere, the K/T boundary will comprise, at best, a few millimetres of impact clay.

The Chicxulub crater in northern Yucatan seems to be the most promising candidate site for the K/T impact. Shocked quartz grains as large as 1 mm across from an 80-m-thick high-energy sediments unit, 50 km from the crater rim, probably show this unit to be made of impact ejecta (A. Hildebrand, University of Arizona). Gravity and magnetic anomalies have been known there for a long time and

were ascribed to buried volcanic lava-flows (andesite), but are possibly due to misinterpreted impact melts. The gravity anomaly looks similar to one at the 210-million-year-old Manicouagan crater in Quebec, but has twice the diameter. Sadly, the warehouse which stored the critical drill-cores from the impact melt of the Chicxulub crater burned down a few years ago and data seemed lost forever. Hildebrand and Boynton, however, tracked down bits and pieces from these cores in private collections, and so discovered the shocked quartz.

Now tektite glass from the Beloc site, Haiti, has been identified¹ in millimetre-sized spherules by S. D'Hondt and H. Sigurdsson (University of Rhode Island). The smectite-coated glass occurs at the base of a 55-cm-thick layer exactly at the K/T boundary in Beloc. Formerly this unit was regarded as the product of volcanic-activity. Most Beloc tektites are spherical, but many splash forms, such as droplets and dumbbells, well-known from other tektites, also occur. The composition of the tektite glass is similar to andesite (60–67 per cent SiO₂), at the low end of the range of known tektites. The important thing about the Beloc glass is that it is completely free of small crystals, and almost free of gas and water. That is unknown for any kind of volcanic glass (eruptions are gas-driven), but diagnostic for tektites. Thus these tektites seem to provide the final riposte to the volcanic-eruption mechanism⁵.

Identical spherules, but completely altered, have been identified higher up in the Beloc K/T layer and at other K/T sites on or close to North America, but these were overlooked as being of local origin. Outside North America, abundant small mafic spherules occur with tiny crystallites of minerals (magnesian-ferrite, clinopyroxene and their alteration products) that are incompatible with continental source rocks (Figs 1 and 2).

Both these 'microkrystites' and the tektites can be produced by a single impact, according to modelling by H. J. Melosh and A. M. Vickery (University of Arizona). Over half of the rock ejected by an impact can remain molten in the plumes and will subsequently cool to form tektites; the higher-energy portion of the plume is vaporized and may condense to form the spherules. The spherules may have an entirely different composition from the tektites and may contain significant amounts of material from the impactor itself. (It is remarkable that the only three known strata with well-defined iridium anomalies – one in the Archaean (D. R. Lowe, Stanford University), the K/T boundary itself, and one in the late Eocene – also include microkrystites.) The K/T microkrystites are enriched in

* Symposium on Impacts and the Terrestrial Environment, AGU Fall meeting, 3–7 December 1990.

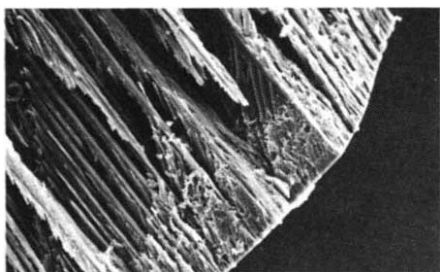


FIG. 2 Detail of a clinopyroxene microkrystite from the KT boundary of DSDP site 577. The mafic glass between the pyroxene has been etched away by sea water. Similar forms are seen in all other known altered microkrystites.

siderophile ('iron-loving') elements, associated with Solar System objects. Models of the dispersion of ejecta clouds also indicate that much of the high-energy material should be deposited on the other side of the globe from the impact. This explains the high levels of iridium and shocked quartz found in the Southern Pacific (at DSDP site 596; F. T. Kyte, University of California at Los Angeles), in New Zealand and on the Kerguelen Plateau (ODP site 738c), if North America was the site of the impact.

Identifying the impact site, whether it is Chicxulub or not, is not the same as unravelling the mechanism that actually caused the extinctions. The original simple picture — a suppression of photosynthesis resulting from a darkening of the skies by ejecta — has been supplanted by a plethora of complex schemes. Oblique impacts may be more efficient in transferring a larger proportion of the impact energy to the atmosphere than vertical collisions (P. H. Schultz, Brown University, Rhode Island), and fast ejecta re-entering the atmosphere may radiate off their energy and ignite global wildfires (Melosh). Shock heating of the atmosphere may lead to NO_x production, and thus to acid rain. Trapped natural gas (clathrates), principally methane, may have been released in large volumes, following disruption of submarine sediments (M. Max, Naval Research Lab, Washington DC). This would contribute to a greenhouse atmosphere, as would the release¹ of 10^{15} moles of CO_2 through impact with carbonate rocks as found at Yucatan. Carbon and oxygen isotope data from southern Spain⁷ suggest a rise in sea surface temperatures of at least 8°C . Identifying the impact site may help to tie down the details of the extinction. □

Jan Smit is at the Institute of Earth Sciences, Free University, PO Box 7161, 1007 MC Amsterdam, The Netherlands.

Positive about water feedback

Robert D. Cess

INTERACTIVE feedback mechanisms constitute the greatest uncertainty in projecting future climate change associated with increasing greenhouse gases. The change can be either amplified (positive feedback) or reduced (negative feedback). One mechanism, involving water vapour, has long been regarded as being positive¹, but recently it was suggested² that it is instead negative. On page 500 of this issue³, however, Rind *et al.* present combined observational and model results which are strongly supportive of the conventional interpretation that water vapour indeed provides positive feedback.

The customary explanation of water-vapour feedback is that if the Earth were to warm, for example as the result of increasing greenhouse gases, then (owing to increased evaporation) the warmer atmosphere should contain more water vapour which, itself a greenhouse gas, would amplify the warming. The most straightforward way of studying this specific feedback is to restrict attention to clear-sky conditions, so as to eliminate the multitude of complexities associated with the additional feedback due to clouds, for which the sign is in doubt⁴. Even with this restriction, in all likelihood water-vapour feedback also includes climate-induced changes in atmospheric lapse rate⁵, the temperature drop with altitude, although little effort has been made at separating this from feedback due solely to water vapour. It is emphasized that lapse-rate feedback is implicit in what is commonly referred to as water-vapour feedback.

To place the study by Rind *et al.*³ in perspective, it is useful to review recent developments in our understanding of water-vapour feedback. A comparison last year¹ of the global-warming simulations from 19 general circulation models (GCMs) of the atmosphere, found excellent agreement among their depictions of positive water-vapour feedback. This is significant with respect to the issue addressed by Lindzen² and by Rind *et al.*; namely, does the increased convection expected in a warmer climate produce positive or negative water-vapour feedback? Lindzen's contention is that convection will cause water-vapour feedback to be negative, because sinking air in regions surrounding deep cumulus clouds will dry the upper troposphere. The point is that the 19 GCMs used several very different parameterizations of convection, so that their agreement over water-vapour feedback is not an artefact of this parameterization.

In an observational evaluation of water-vapour feedback based on satellite infrared and surface-temperature data,

Raval and Ramanathan⁶ used measurements over different oceanic locations as a surrogate for global climate change; their results concur well with the 19 GCMs⁷. One could, of course, argue that for this purpose a change of location is not a proper surrogate for a change of global climate, and that the agreement is coincidental. A recent study, however, makes this unlikely. A.D. Del Genio and coworkers (private communication), using the same GCM as Rind *et al.*³, predict correctly the changes in radiative fluxes with geographical location observed by Raval and Ramanathan. Furthermore, their model is one of the 19 used in the GCM intercomparison¹, so that it is producing comparable water-vapour feedback for both geographical and global climate change, suggesting that changed location, at least for inferring water-vapour feedback, is a useful surrogate for global change.

Still, one might argue that this agreement also is merely a model artefact. But the new contribution by Rind *et al.* to our overall picture of positive water-vapour feedback wouldn't support this. The authors first demonstrate that for CO_2 -induced global warming, their model predicts an increase in the amount of water vapour in the middle and upper troposphere; this should be expected for a GCM that produces positive water-vapour feedback. Their next step is the important one. To validate their model, they consider the effect of seasonal climate changes on middle- and upper-tropospheric water-vapour content, and show, through comparison with satellite-generated data, that the GCM produces realistic results. This reinforces our confidence that the GCM is also producing realistic predictions for the feedback for the CO_2 -induced warming. It is important to recognize that, in this context, seasonal climate change is used as a validation tool rather than as a surrogate for climate change.

Nevertheless, one can argue that this validation constitutes only a necessary, but not a sufficient, test of the GCM. Rind *et al.*³, however, provide a further and intriguing comparison which specifically addresses the effect of enhanced convection on middle- and upper-tropospheric water-vapour levels. They do this by com-

1. Sigurdsson, H. *et al.* *Nature* **349**, 482–487 (1991).
2. Sheehan, P. M. *et al.* *Geol. Soc. Am. Abs. Prog.* **A**, 356 (1990).
3. Grieve, R.A.F. *Nature* **340**, 428–429 (1989).
4. Hildebrand, A. R. & Boynton, W. V. *Science* **248**, 843–847 (1990).
5. Officer, C. B. *et al.* *Nature* **326**, 143–149 (1987).
6. Ross, I. R. *et al.* *Tectonophysics* **155**, 139–168 (1988).
7. Smit, J. *Geologie Mijnb.* **68**, 187–204 (1990).

1. Manabe, S. & Wetherald, R.T. *J. Atmos. Sci.* **24**, 241–259 (1967).
2. Lindzen, R.S. *Bull. Am. Met. Soc.* **71**, 288–299 (1990).
3. Rind, D. *et al.* *Nature* **349**, 500–503 (1991).
4. Cess, R.D. *et al.* *J. geophys. Res.* **95**, 16601–16615 (1990).
5. Wetherald, R.T. & Manabe, S. *J. Atmos. Sci.* **45**, 1397–1415 (1988).
6. Raval, A. & Ramanathan, V. *Nature* **342**, 758 (1989).
7. Cess, R.D. *Nature* **342**, 736–737 (1989).

paring satellite data for atmospheric moisture over the convective western Pacific with that over the largely non-convective eastern Pacific. The data clearly show that convection moistens the middle- and upper-troposphere over the western Pacific relative to the eastern Pacific. This runs right against Lindzen's argument that convection dries the atmosphere.

As with the other comparisons discussed in this article, one can probably add a caveat to the trans-Pacific comparison. But it is difficult to believe that the agreements found in all of these comparisons are collectively coincidental. Instead

they collectively make a compelling case that water-vapour constitutes a positive feedback mechanism, although as previously discussed there is a need to understand better the lapse-rate component of this feedback. Nor does this imply that we necessarily understand other feedback mechanisms. Cloud feedback is a prime example of a crucial feedback that is poorly understood¹. □

Robert D. Cess is in the Institute for Terrestrial and Planetary Atmospheres, State University of New York, Stony Brook, New York 11794, USA.

RNA SPLICING

Alive with DEAD proteins

David A. Wassarman and Joan A. Steitz

IT has been known for many years that splicing of nuclear pre-messenger RNA (pre-mRNA) requires ATP hydrolysis. Yet the phosphates of ATP are incorporated into neither the intermediates nor the products of the splicing reaction, from which it would seem that the catalytic mechanism itself is independent of ATP — as is the case for the self-splicing of group II introns, which proceeds by way of identical reaction intermediates¹. Instead, it has seemed likely that ATP facilitates the assembly of myriad *trans*-acting factors onto the pre-mRNA as a prerequisite to splicing. The resulting 'spliceosome' is known to include four distinct small RNA-protein complexes, as well as several less well characterized protein factors.

Evidence

Many lines of evidence have hinted at dynamic transitions as the spliceosome proceeds through the reaction pathway, ultimately releasing ligated exons and excised intron. On pages 494 and 487 of this issue, Schwer and Guthrie² and Company *et al.*³ describe the characterization of two essential splicing factors from the yeast *Saccharomyces cerevisiae* as members of a family that includes RNA helicases and RNA-dependent ATPases; the factors concerned are PRP16 and PRP22, PRP standing for precursor RNA processing. Finding ATP-driven switch proteins not only sheds light on splicing's ATP requirement but also promises insights into the pathway, mechanism and fidelity of the process.

Yeast genetics has been exploited to identify over 20 proteins which contribute directly to pre-mRNA splicing⁴. Some are integral components of the small nuclear ribonucleoprotein particles (snRNPs), which contain U1, U2, U4/U6 and U5 RNAs; others (designated 'factors') can be identified only in the spliceosome. PRP16 is a factor required *in vitro* for the

second cleavage-ligation step of the splicing reaction (see figure). Schwer and Guthrie² show that stalled spliceosomes pre-formed in a PRP16 immunodepleted extract can be complemented with the purified protein and ATP to yield mature mRNA; PRP16 then dissociates from the spliceosome. PRP22 acts at a late step, release of the spliced mRNA, one that might have not been expected to require a factor. In a heat-inactivated PRP22 extract *in vitro*, Company *et al.*³ find that the mature messenger is retained in the spliceosome.

Sequence analysis of PRP16 and PRP22 — as well as of three other yeast splicing

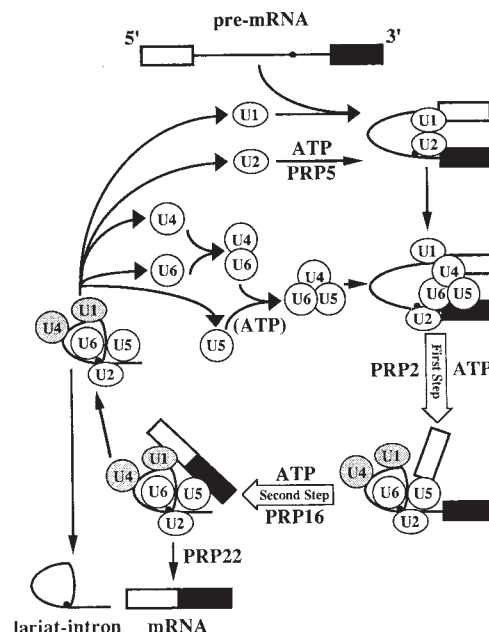
factors, PRP2 (ref. 5), PRP5 (ref. 4) and PRP28 (ref. 6) — shows that they are related to a rapidly expanding family dubbed the 'DEAD box' proteins⁷ (DEAD standing for Asp-Glu-Ala-Asp in the single-letter amino-acid code). Gene products containing the DEAD motif plus six other shared sequence segments come from the whole range of organisms and are implicated in a variety of cellular processes (see table). Among the seven conserved regions are two that make up part of the ATP-binding site in proteins whose structure is known. The eukaryotic protein synthesis initiation factor eIF4A is the prototype⁷ and the member of the DEAD-box family that is the best characterized biochemically⁸. In combination with other initiation factors, it is believed to use the energy generated from ATP hydrolysis to unwind mRNA secondary structure upstream of the initiation codon, thereby facilitating attachment of the 40S ribosome. It should be noted that this unwinding is not of the irreversible type accomplished by covalent base modification, as described for some RNA helicase activities⁹. Only two other members of the DEAD-box family, human p68 (ref. 10) and *Escherichia coli* SrmB (ref. 11), have been biochemically examined and shown to exhibit RNA-dependent ATPase activity (and helicase activity in the case of p68).

As the roster of related proteins grows, the story becomes more complex. In fact, PRP2, 16 and 22 are not literally DEAD-box proteins, but 'DEAH-box'

proteins (H for His). They also share several other sequence features that distinguish them as a novel subclass of the family. In an age when attributing function to a protein based merely on sequence similarity raises few eyebrows, it is particularly gratifying that the biochemical activity of PRP16 has been investigated. Schwer and Guthrie² isolated the protein and directly demonstrated its ability to hydrolyse ATP in the presence of RNA. Whether helicase activity will be found remains to be seen. Admittedly, doing decisive biochemistry on most of the proteins listed in the table is not easy because their exact RNA substrates are unknown.

Speculation

Nonetheless, we can speculate on how a variety of RNA helicases or RNA-dependent ATPases might feed into the splicing pathway. Steps currently known to require factors assigned to the family are shown in the figure. Additional yeast proteins possessing signature sequences of the DEAD (ref. 12) or DEAH (ref. 3) types have already been identified



The roles of DEAD and DEAH proteins in pre-mRNA splicing. The branch point is indicated by ●; shading of the U1 and U4 snRNPs indicates weak interactions with the spliceosome. Note that it is uncertain whether ATP hydrolysis facilitates assembly or dissociation (or both) of the U4/U6 and U5 snRNPs. A more complete model and description of pre-mRNA splicing in yeast can be found in ref. 4.

Protein (ref.)	Organism	Cellular proteins of the DEAD/DEAH family* Cellular function	Biochemical activity
PRP2† (5)	<i>S. cerevisiae</i>	After spliceosome assembly but before first step of splicing ⁴	?
PRP16† (17)	<i>S. cerevisiae</i>	Second step of pre-mRNA splicing, ² suppressor of a branch-point mutation ¹⁷	RNA-dependent ATPase ²
PRP22† (3)	<i>S. cerevisiae</i>	Disassembly of the spliceosome ³	?
PRP5 (4)	<i>S. cerevisiae</i>	Spliceosome assembly ⁴	?
PRP28 (6)	<i>S. cerevisiae</i>	First step of pre-mRNA splicing, suppressed by a mutation in PRP8 (a U5 protein) ⁶	?
Spp81/DED1 (18)	<i>S. cerevisiae</i>	Pre-mRNA splicing, suppressor of a mutation in PRP8 (a U5 protein) ¹⁸	?
MSS116 (7)	<i>S. cerevisiae</i>	Mitochondrial pre-mRNA group II splicing ⁷	?
eIF4A (7)	Mouse	Translation initiation ⁸	ATP-dependent RNA helicase (bidirectional) ⁸
TIF1.2 (7)	<i>S. cerevisiae</i>	Translation initiation, homologous to eIF4A ⁷	?
SrmB (11)	<i>E. coli</i>	50S ribosomal subunit assembly, suppressor of a ribosomal protein mutation ¹¹	RNA-dependent ATPase ¹¹
SPB4 (19)	<i>S. cerevisiae</i>	25S ribosomal RNA maturation ¹⁹	?
vasa (7)	<i>Drosophila</i>	Formation of polar granules and germ cells ⁷	?
PL10 (7)	Mouse	Spermatogenesis ⁷	?
An3 (20)	<i>Xenopus</i>	Unknown, sequence similarities to PL10 ²⁰	?
ME31B (21)	<i>Drosophila</i>	Female germ-line expression ²¹	?
p68 (10)	Human	Cell growth and division ¹⁰	ATP-dependent RNA helicase ¹⁰
RM62 (22)	<i>Drosophila</i>	Unknown, sequence similarities to p68 ²²	?
DbpA (23)	<i>E. coli</i>	Unknown, identified with a p68 probe ²³	?

* Proteins are grouped according to roles in splicing, translation, development or cell growth. The table does not include a number of biochemically uncharacterized viral proteins which can also be assigned to the DEAD-box family. † These comprise the current DEAH-box subfamily.

and may turn out to promote other splicing steps.

First, the initial loading of snRNPs onto the pre-mRNA involves conformational switches in several splicing components. For instance, the U4/U6 snRNP enters the spliceosome complexed with U5, and there is abundant evidence that the assembly/disassembly of this tri-snRNP particle is responsive to ATP¹. Juxtaposition of the 5' splice site and branch point achieved by binding of the U1 and U2 snRNPs may entail active unfolding and refolding of the intron, especially if it is very large. Binding of the U2 snRNP to the pre-mRNA is known to require ATP. PRP5 (DEAD) probably participates in one of these early events, because yeast spliceosomes do not form in its absence⁴.

Second, there are three well-documented RNA-RNA base-pairing interactions which must be dissolved during splicing: U1 with the 5' splice site; U2 with the branch-point region; and U4 with U6. The U4 and U6 snRNAs exist in a single stable particle held together by substantial intermolecular helices termed stem I (8 base pairs) and stem II (16 base pairs)¹³. After the U4/5/6 snRNP complex joins the spliceosome, but before the first cleavage, association of the U4 particle is destabilized. PRP28 (DEAD) is an excellent candidate helicase⁶ because it interacts genetically with PRP24, a U6 snRNP protein (K. W. Shannon and C. Guthrie, personal communication). ATPases could likewise switch snRNPs between alternative forms, thereby changing the accessibility of snRNA sequences as

splicing progresses. Accordingly, ATP is known to stimulate deoxyoligonucleotide-directed RNase H cleavage of U2 sequences complementary to the intron branch point, of U6 sequences involved in stem I base pairing, or of U4 sequences involved in stem II base pairing¹⁴. Recently a fourth base-pairing interaction, between the U2 and U6 snRNPs, was revealed by psoralen crosslinking¹⁵. It foreshadows the probable discovery of additional snRNP-snRNP contacts, each a likely target of ATPase/helicase activity.

Third, spliceosome disassembly not only involves release of the mature mRNA for transport to the cytoplasm and of the excised intron leading to its rapid degradation, but it also enables snRNPs to recycle into new rounds of splicing. It has been proposed that U6 RNA has a catalytic role in splicing and that rejoining of the U4 and U6 snRNPs after spliceosome dissociation sequesters U6 in an inactive form¹. But, in standard *in vitro* splicing systems, snRNPs are not efficiently released from the intron lariat, suggesting that a component or process is missing. PRP22 (DEAH)³ may be only the first factor assigned to a late phase of the splicing cycle.

Mechanistic parallels between the nuclear pre-mRNA splicing apparatus and the cytoplasmic translational machinery are thus becoming more and more obvious. Both involve the ordered assembly of ribonucleoprotein subunits onto an RNA substrate. Conformational changes induced by ATP or GTP hydrolysis mediate the sequential binding and

ejection of factors, enforcing an orderly progression from one step to the next in the pathway. Putative RNA helicases, as well as proteins carrying zinc-finger, arginine-rich or RNP-consensus motifs, play key roles in both processes¹.

But perhaps the most exciting insights provided by a comparison of spliceosome and ribosome function are to do with the problems of fidelity and regulation. During polypeptide-chain elongation, fidelity is achieved by a kinetic proof-reading mechanism that discards an incorrect transfer RNA/EFTu/GTP complex from the ribosomal A site with concomitant GTP hydrolysis¹⁶. PRP16 (DEAH) was in fact first identified as a dominant suppressor of a branch-point mutation in the pre-mRNA¹⁷; that is, a mutant PRP16 is able to decrease the fidelity normally exercised in using a branch-point region, allowing a mutant substrate to be spliced.

The PRP16 mutant protein can now be tested directly for an alteration in its rate of ATP hydrolysis, which would support the idea¹⁷ that comparable energy-consuming proof-reading mechanisms have evolved for the spliceosome. It remains to be seen how many DEAD- or DEAH-box splicing factors can mutate to suppress pre-mRNA mutations. Will mutations in some of their mammalian analogues (when discovered) alter the recognition pattern of alternative splice sites, hinting at contributions to the regulation of splicing? How many family members really are helicases? How many are needed simply to couple the complexity of spliceosome dynamics to ATP hydrolysis? We must wait and see. □

David A. Wassarman and Joan A. Steitz are in the Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA.

- Guthrie, C. *Am. Zool.* **29**, 557–567 (1989).
- Schwer, B. & Guthrie, C. *Nature* **349**, 494–499 (1991).
- Company, M., Arenas, J. & Abelson, J. *Nature* **349**, 487–493 (1991).
- Ruby, S.W. & Abelson, J. *Trends Genet.* (in the press).
- Chen, J.-H. & Lin, R.-J. *Nucleic Acids Res.* **18**, 6447 (1990).
- Strauss, E.J. & Guthrie, C. *Genes Dev.* (in the press).
- Linder, P. *et al.* *Nature* **337**, 121–122 (1989).
- Rozen, F. *et al.* *Molec. cell. Biol.* **10**, 1134 (1990).
- Bass, B.L. & Weintraub, H. *Cell* **55**, 1089–1098 (1988).
- Hirling, H. *et al.* *Nature* **339**, 562–564 (1989).
- Nishi, K. *et al.* *Nature* **336**, 496–498 (1988).
- Chang, T.-H., Arenas, J. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1571–1575 (1990).
- Brow, D.A. & Guthrie, C. *Nature* **334**, 213–218 (1988).
- Black, D.L. & Steitz, J.A. *Cell* **46**, 697–704 (1986).
- Hausner, T.-P., Giglio, L.M. & Weiner, A.M. *Genes Dev.* **4**, 2146–2156 (1990).
- Thompson, R.C. *Trends biochem. Sci.* **13**, 91–93 (1988).
- Burgess, S., Couto, J.R. & Guthrie, C. *Cell* **60**, 705–717 (1990).
- Jamieson, D.J. *et al.* *Nature* (in the press).
- Sachs, A.B. & Davis, R.W. *Science* **247**, 1077–1079 (1990).
- Gururajan, R. *et al.* *Nature* (in the press).
- deValoir, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Dorer, D.R., Christensen, A.C. & Johnson, D.H. *Nucleic Acids Res.* **18**, 5489–5494 (1990).
- Iggo, R. *et al.* *Nucleic Acids Res.* **18**, 5413 (1990).

The case of the curious bill

John R. Krebs

THE evolution of complex traits is a classic problem in natural selection. How can a character such as the vertebrate eye, which relies on the coordinated functioning of many separate parts, evolve by the gradual accretion of improvements? Surely, the argument goes, intermediates would be of no selective advantage because the character 'works' only if all its parts function together. The usual answer to this point, expressed forcefully and clearly by Richard Dawkins¹, relies on advocacy rather than direct evidence: "Without an eye you are totally blind. With half an eye you may at least be able to detect the general direction of a predator's movement And this may make all the difference between life and death". For this reason, the results reported on page 519 of this issue² by Benkman and Lindholm are of general interest. Their experiments show how a complex and improbable trait, the crossed mandibles of the crossbill, can have a potential selective benefit that increases as the expression of the trait increases.

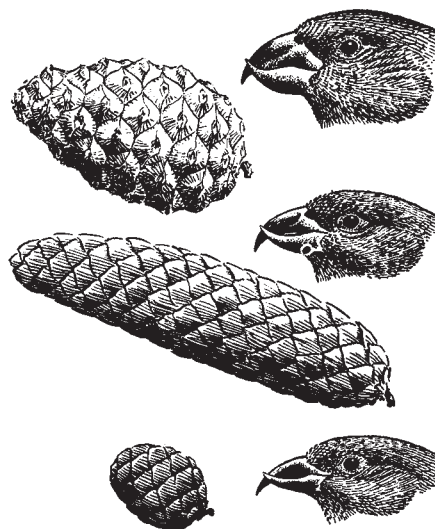
Crossbills have remarkable beaks: the lower mandible twists either to the right or left of, and crosses over, the upper one (see figure). According to legend, the crossbill's mandibles acquired their shape as a result of the birds trying to pull the nails from the hand of Jesus on the cross, whilst the red colour of the male was said to have arisen from the stain of Christ's blood³. A more utilitarian, darwinian view is that the beak is an adaptation for removing seeds from conifer cones, allowing crossbills access to parts of the cone that other birds cannot reach. Newton⁴ gives the following description of how this is done.

A bird first wrenches off a cone in its bill, carries it to some firm, horizontal branch, and clamps it between one of its feet and the perch. It then works the tip of its beak behind one of the scales, while the cone is held so that it points forwards and slightly to one side. A bird with the lower mandible deflected to the right holds the cone in its right foot and vice versa. The result is that, whether the bill crosses to the right or left, the tip of the lower mandible is always brought to bear on the cone itself and the upper one on the inside of one of the scales. The lower mandible is then moved sideways towards the body of the cone, so that the scale is raised by the tip of the upper one. The seed, once released, is scooped out by the protrusible tongue.

It is widely accepted that the shape of birds' bills reflects selection for increased feeding efficiency, and as long ago as 1944 David Lack⁵ pointed out that the three species of crossbill in Europe have beaks that match their preferred food (see figure). The parrot crossbill (*Loxia pytyop-*

sittacus) has the largest and most heavily built beak and it feeds on tough cones of *Pinus*, the intermediate common crossbill (*L. curvirostra*) feeds on slightly less hard cones of *Picea*, whilst the slender bill of the two-barred crossbill (*L. leucoptera*) is used for feeding on the small, soft cones of *Larix*. Long-term studies of Darwin's finches have shown that natural selection can act to produce rapid changes in beak shape or size within a species in relation to year-to-year changes in seed hardness⁶. Even within the lifetime of an individual, subtle, reversible, seasonal changes in beak shape and size may be related to changes in feeding niche⁷.

In their experiment, Benkman and Lindholm used nail clippers to trim the



Heads of the three European crossbill species, and the main cones eaten. *Top*, parrot crossbill and pine cone. *Centre*, common crossbill and spruce cone. *Bottom*, two-barred crossbill and larch cone. (Reproduced from *Finches* by Ian Newton, Collins, 1972.)

crossed ends of the bills of crossbills. This manipulation, akin to clipping toenails, causes no serious damage and, further, the crossed mandibles gradually regrow over a period of weeks. This mimics the process of growth of the young crossbill's beak. Nestlings have uncrossed bills: crossing over starts at about the age of 27 days (one week after fledging) and is complete by about 45 days, when youngsters are able to feed effectively on cones.

Benkman and Lindholm measured the efficiency with which a subspecies of common crossbill (*L. curvirostra minor*) extracted seed from hemlock (*Tsuga heterophylla*) cones both immediately after bill docking and during the four-week period of gradual regrowth of the crossed mandibles. They found that the time taken to extract a seed from closed hemlock cones decreased gradually with

increasing regrowth of the crossed mandibles. In other words, performance improved progressively with increasing expression of the trait. This was not due simply to the birds' learning to cope with trimmed bills, because the time required to extract a seed from open cones (cones open as they dry out) did not increase in the same way with regrowth of the mandibles. Nor was the effect due to a general impairment of foraging ability, illustrated by the fact that bill trimming had no effect on the time required to handle and consume a seed once it had been extracted. The birds with trimmed bills were similar in their performance to another species, the pine siskin (*Carduelis pinus*), which forages on open pine cones but is relatively unsuccessful at extracting seeds from closed cones.

Does this experiment provide a satisfactory demonstration that there could in principle be progressive selection for the evolution of a complex trait? The answer is only partially in the affirmative. The experiment manipulated only one aspect of the trait. As well as a crossed bill, successful extraction of seeds from closed cones requires other morphological traits such as asymmetric jaw muscles (the asymmetry has to match the 'handedness' of the bill), an elongated, cartilage-containing tongue and a flexible jaw hinge. Further, these morphological traits have to be integrated with the appropriate motor patterns. Benkman and Lindholm show that if the crossbill has all the other features required for cone opening, it is more efficient at extracting seeds with a crossed bill. This does not answer the question of how bill, muscle, tongue and behaviour came together during evolution, nor does it indicate whether morphological evolution is driven by behaviour or the other way round.

Nevertheless the study offers a glimpse of a possible model system in which to study such questions. Bird fanciers have produced hybrids between crossbills and other finches such as the greenfinch (*Carduelis chloris*) with less specialized bills⁴. Perhaps an analysis of the genetics, morphology and behaviour of these hybrids would provide an avenue along which to explore further the evolution of a complex trait. □

John R. Krebs is in the Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Dawkins, R. *The Blind Watchmaker* (Longman, London, 1986).
2. Benkman, C.W. & Lindholm, A.K. *Nature* **349**, 519–520 (1991).
3. Nethersole-Thompson, D. *Pine Crossbills* (T. & A.D. Poyser, Berkhamsted, 1975).
4. Newton, I. *Finches* (Collins, London, 1972).
5. Lack, D. *Ibis* **86**, 552–553 (1944).
6. Grant, P.R. *Ecology and Evolution of Darwin's Finches* (Princeton, Univ. Press, 1986).
7. Gosler, A.G. *Ibis* **129**, 451–476 (1987).

Just a light squeeze

R. S. Pease

LASER radiation has been used to compress matter to a record density in experiments detailed at a recent conference* on plasma physics. Part of the drive to generate nuclear fusion using 'inertial confinement', the experiments (S. Nakai, Osaka Univ.) used 8–10 kilojoule pulses of focused, short-wavelength laser light to compress small carbon–deuterium–tritium capsules to a density of 600 tons per cubic metre, about three times that reported from other laboratories (R. McCrory, Rochester Univ.).

Expressed in terms of the number of electrons per unit volume, the peak density at 10^{32} m^{-3} reported from Osaka is comparable to that in our nearest natural fusion reactor, the Sun's core. The measured temperature, on the other hand, is about a fifth that in the Sun's core. One consequence is that the characteristic 'Fermi' energy of such compressed plasma (about 1,000 eV) exceeds the thermal energy (300 eV), so that a regime in plasma physics new to the laboratory, in which the electrons are degenerate — most in the lowest energy states — may be created.

The measured densities, obtained by monitoring the number of secondary reactions of the fusion products with heavier nuclei present in the laser pellets (which depends on the density and radius) and an estimate of the mass of matter compressed, are accurate to within a factor of two or so. Nevertheless, they agree with that calculated on the assumption that the laser pulses implode the pellets spherically symmetrically. But the fusion neutron yields tell a different story. For radial compression ratios more than about 20, the yield is lower than that calculated — by three orders of magnitude at the highest compression. This corresponds (roughly) to the ion temperature being three times lower than expected. It could be that non-uniformity in the implosion results in some cold material being mixed into the innermost hot spark (Nakai).

The compression achievable by these techniques is limited by the Rayleigh–Taylor instability, first investigated by Lord Rayleigh (*Scientific Papers* Vol. 2, 200, Cambridge Univ. Press; 1900), who was interested in the structure of cirrus clouds in the upper atmosphere. The problem is that the inner imploding region, accelerating inwards at around 10^{14} m s^{-2} , is more dense than the laser-heated outer layer that drives the implosion. Rayleigh's studies show that in these circumstances, any imperfection will quickly grow, the growth rate for an acceleration g

being given as $\sqrt{2\pi g/L}$, where L is the wavelength of the initial imperfection when the boundary between the two layers is sharp, or the density length scale of the boundary when it is diffuse (whichever is the longer dominates). Although the use of time-varying laser intensities, multiply layered targets and so on allows the designer to lessen the instabilities to a degree, the Rayleigh–Taylor instability can also occur at the centre of the pellet where imploding material is brought to a halt by a less dense, hot plasma. The potentially infinite growth rate that results from the approach of L to zero on these scales is avoided in ordinary fluids by the intervention of surface tension.

But surface tension seems to have little direct effect on the imploding plasma (H. Hora *et al. Trans. Plasma Sci. IEEE* 17, 284; 1989) and instead the instabilities may be stabilized by kinetic effects in the ablating outer layer. Computer calculations suggest that perturbations with wavelengths less than about 100 μm (depending on the velocity of the ablated ions) might be stabilized in this way (H. Takabe, Osaka Univ.) and that growth rates at longer wavelengths are reduced. Experiments at Lawrence–Livermore indicate that the growth of 100- μm -wavelength disturbances is reduced by around a factor of two (E. Storm). A UK group, using the Rutherford–Appleton Laboratory's central laser facility, report (M. Desselberger *et al. Phys. Rev. Lett.* 65, 2977–3000; 1990) Rayleigh–Taylor instabilities growing at L values of 30 μm . They did not show that ablation stabilization occurred.

Thus much of the current effort in research into inertial-confinement fusion is in minimizing the initial imperfections and in mitigating the Rayleigh–Taylor growth rates rather than eliminating them. The advance reported by the Osaka group, for example, depends on a new method for manufacturing uniform, spherically symmetric targets, the carbon–deuterium–tritium plastic shells being produced to a thickness uniform to 1 per cent. And it is now routine to use special optics for the irradiating laser beams, tailored to give a uniform distribution of intensity. Although the principal aim is to develop fusion reactors, perhaps the greatest interest in the new results lies in the prospect of studying the physics of the densest states of matter in the laboratory. □

R. S. Pease, formerly Fusion Programme Director for the UK Atomic Energy Authority, is at The Poplars, West Ilsley, Newbury RG16 0AW, UK.

Elevated circles

SOONER or later, the new ceramic high-temperature superconductors should find their way into power cables, which could then carry thousands of amps per square millimetre. Daedalus now points out a novel consequence of stringing such a current-dense cable across the landscape. If it carried direct current eastwards, the simple motor force of the Earth's magnetic field could levitate the cable.

At first sight, this seems a splendid way of stringing a power line between two points without expensive supporting pylons. Hundreds or even thousands of kilometres might be spanned by one magnetically levitated loop reaching far up into the ionosphere. Indeed, this is an ideal way of supporting and powering TV relay stations, space telescopes, Earth-scan cameras, and so on. But Daedalus is taking the idea to its logical conclusion. He is inventing a 'space halo' — a complete loop of cable, carrying an endlessly flowing superconductive current, magnetically levitated right around the Earth above the Equator.

The Earth's magnetic field decreases with height. So will its lift on the halo, which will therefore reach a stable equilibrium altitude. The latitudinal stability of the halo is not so obvious. A set of current-sharing subloops deviating north and south may be needed to keep it safely equatorial. Such a divided, braid-like structure would still be kept taut and expanded by the strong self-repulsion of its ring current.

This bold experiment is clearly worth trying. But simply laying 40,000 kilometres of superconducting cable loosely round the Equator and starting a current in it, is obviously impractical. Daedalus proposes instead to lay it in a meandering path around the Antarctic ice cap. (The polar cold would simplify refrigeration of the superconductor, too.) As the current was built up, the meanders would levitate to form a ring of arches. Adjacent arches would be merged by releasing the clamps between them, and more current fed in, till the cable was held down at only two or three points, arching between them up and out far into space. On releasing the final clamps, the halo would expand under its self-repulsion and swing through space to its equilibrium equatorial position.

A space-halo should be most useful at an 'auroral' height of 50–100 km, too high for aircraft and too low for satellites. The tenuous atmosphere up there would slowly damp its initial oscillations and drag it gradually up to equatorial geosynchronous speed. Yet in such thin air, simple radiation shields to protect it from sunlight and earthlight while exposing it to the cold of space in other directions, could keep it safely cold and superconducting. It would stay up indefinitely.

David Jones

*13th IAEA Conference on Plasma Physics and Controlled Nuclear Fusion Research, Washington DC, 1–6 October 1990.

Sun and water in the greenhouse

SIR — In their review article on the greenhouse effect¹, Hansen and Lacis stated that measurements of changes in solar diameter limit solar brightness variations to a few tenths of a per cent, citing an article² by one of us (R.I.G.) in support of this estimate. But the article does not support their statement: it actually states that solar diameter changes would constrain solar brightness variations only to a limit of 10%.

The second source for limits on solar changes cited by Hansen and Lacis is a calculation based on the implausible assumption that magnetic fields of the order of a million gauss exist at the base of the solar convective zone³. This assumption is unnecessary as fields of 3–5 kilogauss are sufficient to explain the strength of the fields erupting on the Sun's surface; and it is also extremely improbable, as a million-gauss field at the base of the convective zone would create violently disruptive buoyancy forces and magnetic stresses in the Sun's interior. There is no observational or theoretical evidence to preclude changes of up to 1–2% in the Sun's brightness over century timescales, or linear trends of the order of 0.1% per decade.

Hansen and Lacis note that the climate impact of such changes is comparable to the impact of the greenhouse effect. They also note, contrary to the thrust of their main conclusion, that "measurements of solar irradiance during the next decade or two are crucial for determination of long-term climate trends".

RONALD I. GILLILAND

*Space Telescope Science Institute,
Baltimore,
Maryland 21218, USA*

ROBERT JASTROW

*Department of Earth Sciences,
Fairchild Science Center,
Dartmouth College,
Hanover,
New Hampshire 03755, USA*

EUGENE PARKER

*Enrico Fermi Institute,
University of Chicago,
Chicago,
Illinois 60637, USA*

SIR — Hansen and Lacis¹ criticize my arguments² concerning the possible role of cumulus convection in providing a strong negative feedback. Contrary to their suggestion, our calculations do include "moisture detrainment, cirrus anvils, large-scale dynamics and other processes". We recognized that water vapour in the upper troposphere (above about 7 km) is, molecule for molecule, much more important to the temperature of the Earth than water vapour near the surface. Thus, reductions in water vapour at these levels

might more than compensate for increases near the ground.

The evidence cited by Hansen and Lacis against our argument is irrelevant. Oort and Rasmussen⁵ analysed data for humidity below about 6 km; Raval and Ramanathan⁶ looked at vertically integrated water vapour which is dominated by water vapour near the ground. Neither relate to upper-level water vapour. Even Hansen *et al.* state in a recently published letter⁷ "it has been argued that the upper troposphere could become drier as the lower atmosphere becomes moister. Thus, reliable evaluation of the water vapour feedback requires monitoring of water vapour, especially in the upper troposphere, in conjunction with climate changes."

RICHARD S. LINDZEN

*Massachusetts Institute of Technology,
Cambridge,
Massachusetts 02139, USA*

HANSEN AND LACIS REPLY — Both Gilliland *et al.* and Lindzen in their arguments fail to see the forest for the trees.

Lindzen argues that moist (cumulus) convection, expected to increase with global warming, dries the upper troposphere, thus causing a negative climate feedback. Global climate models, some incorporating cumulus subsidence drying, all yield the opposite result: greenhouse warming increases the absolute humidity in the upper (and lower) troposphere. Lindzen⁴ does not present calculations including moisture detrainment into cirrus anvils, large-scale dynamics and other important processes. But his arguments create the impression of a complex disagreement among theorists.

A simple acid test is provided by the real world: does the summer increase of moist convection dry the upper troposphere? Radiosonde data indicate summertime moistening, but Lindzen dismisses these as inaccurate. Precise profiles of water vapour have been measured by satellite⁸: these show summertime moistening of the upper troposphere at all latitudes. Measurements of the Earth radiation budget⁹, which take into account the effect of the entire atmospheric column, independently show that greenhouse warming is largest in regions of moist convection, precisely where Lindzen's 'theory' predicts a minimum. Our call for global monitoring of all major climate forcings and feedbacks⁷ is intended to quantify causes of climate change. We did not intend to provide evidence for a negative water vapour feedback which, in fact, is excluded by existing data.

Our discussion of solar variability referred to a report¹ arguing that reductions of greenhouse-gas emissions "could

turn out to be unnecessary or even harmful" because solar variations may cancel greenhouse warming. We pointed out that one cannot simultaneously assume high climate sensitivity to solar change and low sensitivity to greenhouse-gas changes. The simplest way to look at mechanisms for climate change is to compare the forcings. This procedure shows that the Sun's irradiance must decrease by 2% to cancel the anthropogenic greenhouse forcing that will exist in the first half of next century. There is no evidence for such a large long-term solar decline. Satellite data show a net change much less than 0.1% since 1978 (ref. 10); observations corrected for atmospheric transparency changes yield an upper limit of a few tenths of a per cent this century¹¹; some solar-type stars vary by several tenths of a per cent¹².

The only other evidence we cite on solar variability is the upper limit on solar diameter changes during the past two centuries. These data and Sofia's calculations (we do not use Gilliland's) place an upper limit of a few tenths of a per cent on solar variations^{13,14}. We agree that the result depends on the solar model used, and thus diameter measurements alone add little to the above limits on solar variability. But these data certainly provide no evidence for large changes in solar irradiance.

We stated that "the possibility of a decline of solar irradiance by 2% . . . has not been ruled out categorically". But there is no evidence for such an impending large decline. Thus our conclusion, "Given available scientific evidence, it would be foolish to base greenhouse policy on the hope that solar variability will somehow counteract greenhouse warming", seems to hit the nail on the head.

JAMES E. HANSEN

ANDREW A. LACIS

*NASA Goddard Institute for
Space Studies,
2880 Broadway, New York,
New York 10025, USA*

1. Hansen, J. E. & Lacis, A. A. *Nature* **346**, 713–719 (1990).
2. Gilliland, R. I. *Astrophys. J.* **253**, 399–405 (1982).
3. Endal, A. S., Sofia, S. & Twigg, L. W. *Astrophys. J.* **290**, 748–757 (1985).
4. Lindzen, R. S. *Bull. Met. Soc. Am.* **71**, 288–289 (1990).
5. Oort, A. H. & Rasmussen, E. M. *Prof. Pap. NOAA* **14**, 323 (1971).
6. Raval, A. & Ramanathan, V. *Nature* **342**, 758–761 (1989).
7. Hansen, J., Rossow, W. & Fung, I. *Issues Sci. Tech.* **7**, no. 1, 62–69 (1990).
8. Rind, D., Chiou, E.-W., Chu, W., Larsen, J., Oltmans, S., Lerner, J., McCormick, M. P. & McMaster, L. *Nature* **349**, 500–503 (1991).
9. Seitz, F., Jastrow, R. & Nierenberg, W. *Scientific Perspectives on the Greenhouse Problem* (Marshall Institute, Washington, DC, 1989).
10. Willson, R. C. & Hudson, H. S. *Nature* **332**, 810–812 (1988).
11. Hoyt, D. V. *Rev. Geophys. Space Phys.* **17**, 429–458 (1979).
12. Radick, R. R., Lockwood, G. W. & Baliunas, S. L. *Science* **247**, 39–44 (1990).
13. Sofia, S. *Geophys. Monogr. Ser.* **29**, 202–206 (1984).
14. Endal, A. S., Sofia, S. & Twigg, L. W. *Astrophys. J.* **290**, 748–757 (1985).

Greenhouse budgets

SIR—Hammond *et al.* suggest¹ that for political purposes the relative importance of greenhouse gas emissions from various countries should be based on the observed rate of increase of atmospheric concentration rather than on calculated contributions to future concentrations as suggested by the IPCC (refs 2–4). The advantages alleged by Hammond *et al.* are a lack of dependence on inaccessible models and the removal of arbitrary choices of the time horizon considered. But their choice of a one-year time horizon is itself arbitrary and the use of observed atmospheric increases to assess relative source contributions is equivalent to assuming an effectively constant airborne fraction. This assumption breaks down when increases depart from exponential growth, so in our view is inappropriate for planning possible reductions.

As an example, the currently observed increase in CH₄ reflects continually increasing sources. If emissions stabilized then the concentration would also soon stabilize. The approach of Hammond *et al.* would then treat CH₄ emissions as harmless even though present emissions would keep atmospheric CH₄ levels (and the consequent radiative forcing) at more than double the natural pre-industrial level.

If fairness is desired, then we can see no alternative to modelling relations between sources and concentrations. Even highly sophisticated models can be parameterized so that model calculations can be performed with personal computers, spreadsheets or by hand (see ref. 4). One change to the IPCC approach that we do suggest is the use of CH₄ rather than CO₂ as a reference case. CH₄ has the advantages of having a more linear relation between concentration and radiative forcing and a response that is well characterized by a single lifetime. Also, this lifetime is relatively short so that the integrated effect rapidly ceases to change as the time horizon is increased beyond a few decades. By ceasing to use CO₂ as a reference case, the special problems of CO₂ only appear when discussing CO₂ and are not factored into all other comparisons.

I. G. ENTING

Division of Atmospheric Research,
CSIRO,
Private Bag 1, Mordialloc,
Victoria, 3195 Australia

H. RODHE

Department of Meteorology,
Stockholm University,
Stockholm S-10691,
Sweden

SIR—The proposal of Hammond *et al.*¹ to account for contributions of greenhouse gases to the atmosphere has a major virtue

that the authors have failed to claim in that it emphasizes the rate of warming. Although Hammond *et al.* did not mention this point, I believe that it is crucial. These authors suggest that each country's annual contribution to increases in the greenhouse effect can be determined by weighting that country's emission of each gas by the global atmospheric increase in that gas divided by the total anthropogenic source of that gas (the so-called airborne fraction), and then multiplying by the infrared heating effectiveness of that gas. I will refer to these as greenhouse forcing contributions (GFCs). Because a year is short compared to the lifetimes of most of the greenhouse gases, this is essentially an instantaneous measure of change of radiative forcing, as Hammond *et al.* note (though not a direct measure of warming).

It is the rate of climate change, rather than the eventual magnitude of the total anthropogenic change, that will be most troublesome. Warming the Earth by a few kelvin over the course of several millennia would change the planet's ecosystems, although probably not catastrophically. The evolution of human systems and civilization forced by anthropogenic climate change would then become small compared to most other developmental pressures, which is probably a reasonable goal.

Some policy makers and laypeople suggest that attempts to slow greenhouse climate change will cause great economic and social upheaval, and that if this delays climatic disaster by only a few decades, it is not worth the bother. The investment and development timescales for agricultural and energy systems are about 30 years. If we slow climate change down to the level that significant change occurs over at least two or three such periods, rather than one, I believe we will have done something useful. We will also have provided a small reprieve for the ecosystems.

Use of the scheme of Hammond *et al.* presents an issue that is not yet a problem, but may become one. Since there are feedbacks between climate change and both CH₄ and CO₂, there may eventually be a time when the increases in the atmospheric content of these gases exceed direct anthropogenic output. In fact, such feedbacks can lead to changes in the atmospheric concentration of greenhouse gases in the absence of any anthropogenic sources (and probably have in the past). This is distinct from the claim by Enting and Rodhe (left) that concentrations would soon stabilize if emissions stabilize. Some nations might strongly object to being credited with greater emissions than they actually release, especially as the initial cause of that atmospheric increase might be the result of warming from different greenhouse gases released by other nations in the distant past. The emission weighting coefficients should

probably not be allowed to exceed one.

As another small amendment to the proposal of Hammond *et al.*, not only should scaling of emissions be updated each year, but so should the infrared heating effectiveness coefficients be updated. These will change due to increasing knowledge, the nonlinear dependence of absorption on concentration of some greenhouse gases, and the changing concentrations of other gases with overlapping absorption bands.

Global warming potentials (GWPs), such as those proposed by the IPCC², treat all greenhouse forcing, regardless of when it occurs, as if it were equal. Further, such potentials are scenario-dependent and require predictions of the future mix of greenhouse gases³. There is a fundamental error in using the proposed GWPs in that it masks responsibility for the rate of climate change. I believe that short-term (annual) GFCs, with attributed contributions limited to actual emissions, are a better choice for determining national responsibility for increasing greenhouse gases.

MARK DAVID HANDEL

Center for Meteorology and Physical
Oceanography,
Massachusetts Institute of Technology,
Cambridge,
Massachusetts 02139, USA

HAMMOND *ET AL.* REPLY—Under the IPCC scheme, the impact of emitting a kilogram of a greenhouse gas is given by a GWP, which can be represented as $I_0 L(T)$, where I_0 is the current infrared heating efficiency and $L(T)$ is an effective lifetime, including the concentration-dependent variation of infrared heating efficiency, and is parameterized by a chosen integration period or time horizon T . A similar quantity, which following Handel we will call the GFC, is represented in our scheme by $I_0 A_1$, where A_1 is the empirically determined airborne fraction.

By comparison, the airborne fraction can also be thought of as an effective lifetime, but one which is related to the current and past behaviour of the atmosphere rather than to its projected future behaviour. A GFC is not useful for predicting the future, but we believe that it is very useful for representing the relative impact of different gases in a manner suitable for diplomatic agreements.

We disagree with Enting and Rodhe that a one-year time horizon is arbitrary—

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

it is, in effect, an instantaneous measure. We also disagree that the airborne fraction is an inappropriate parameter: given the current structure of emissions among industrial and developed countries, there is unfortunately virtually no likelihood of stabilizing — let alone reducing atmospheric concentrations for any major greenhouse gas in the next 20–25 years.

ALLEN L. HAMMOND

ERIC RODENBURG

World Resources Institute,
1709 New York Avenue NW,
Washington, DC 20006, USA

WILLIAM MOOMAW

Center for Environmental Management,
Tufts University,
Medford, Massachusetts 02155, USA

1. Hammond, A. L., Rodenburg, E. & Moomaw, W. *Nature* **347**, 705–706 (1990).
2. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change — The IPCC Scientific Assessment* (Cambridge University Press, 1990).
3. Lashof, D. A. & Ahuja, D. R. *Nature* **344**, 529–531 (1990).
4. Maier-Reimer, E. & Hasselmann, K. *Clim. Dyn.* **2**, 63–90 (1987).
5. Victor, D. *Nature* **347**, 431 (1990).

Testing times

SIR—Although I do not dispute the general case made by Blinkhorn and Johnson¹ against the spurious uses to which personality testing is all too often put in the business world, let us not condemn the good along with the bad. The chief abuse which Blinkhorn and Johnson rightly castigate lies in the simultaneous application of large numbers of scales to small samples of people, followed by statistically unsound interpretation of the many coefficients of correlation with performance indices thus obtained.

But there is another, sounder, tradition in scientific research on personality which can yield useful practical, including business, applications. I refer to studies (and there is a substantial corpus of them) in which fewer, broader and statistically independent personality traits, or dimensions², have been shown to correlate reliably with experimental measures of behaviour in ways predicted by relevant theory. To give just two examples: if a job requires prolonged perceptual vigilance under boring circumstances (such as proof-reading), it is better (other things being equal) to choose someone of inverted personality³; and if it requires persistence in spite of repeated rejection (such as selling insurance), you would do well to pick people with an optimistic attributional style⁴.

Before your readers accept too readily Blinkhorn and Johnson's overly pessimistic conclusion that there is "precious little evidence that even the best personality tests predict job performance", they should read at least the second of these reports. This demonstrates a highly successful application of personality testing in the US insurance industry⁵, with

results that we, in an as-yet-unpublished study, have been able substantially to replicate in the UK.

JEFFREY A. GRAY

Department of Psychology,
Institute of Psychiatry,
De Crespigny Park,
Denmark Hill,
London SE5 8AF, UK

1. Blinkhorn, S. & Johnson, C. *Nature* **348**, 671–672 (1990).
2. Eysenck, H. J. (Ed.) *A Model For Personality* (Springer, New York, 1981).
3. Seligman, M. E. P. & Schulman, P. *J. Personality soc. Psychol.* **50**, 832–838 (1987).

SIR—Blinkhorn and Johnson could well be correct in asserting that there is "precious little evidence that even the best personality tests predict job performance". Such a condemnation may however be unjustified.

Astonishingly, the authors make no reference to the context in which the test is used. A very basic lesson in psychometrics is that the 'validity' of a test depends critically on the base rate of the characteristic in the population under test and the selection ratio relating applicants to posts. In the present context, the base rate refers to

Percentage accuracy of a hypothetical 'low validity' test in classifying applicants as suitable or unsuitable			
Test classification		Actual suitability	
		+	–
+	+	25	2.5
	–	75	97.5

the percentage of applicants who would actually be suitable; to the extent that this deviates from 50%, the usefulness of any test procedure in reducing selection errors is impaired. The selection ratio refers to the number of applicants for each available position. Where this ratio is high even tests of very low predictive power can make a contribution. As an extreme example, consider the hypothetical test in the table which inappropriately rejects 75% of suitable applicants. Overall, in classifying a pool of applicants of whom 50% are suitable it has a 38.75% error rate — hardly appealing as a valid test.

Consider now the use of this test to appoint 20 individuals from a pool of 160, of whom only 50% are actually suitable. Without using the test, and selecting randomly, we would expect 50% of the appointees to be suitable — a total of only ten. However, using even this 'invalid' test can be expected to increase the number of suitable appointees to a little over 18 out of the 20. Of course a total of 60 suitable applicants have also been unjustly rejected by the test — but one might cynically expect this to be of little concern to the organization. Such arguments can be extended beyond categorical (suitable/unsuitable) variables to continuous ones, and can of course take account also of the

utilities (in a bayesian sense) associated with different outcomes.

R. GLYNN OWENS

Department of Clinical Psychology,
University of Liverpool,
PO Box 147, Liverpool L69 3BX, UK

BLINKHORN AND JOHNSON REPLY—It is certainly the case that even tests with rather low predictive validity may have practical use when selection ratios and/or base rates take on appropriate values. This much has been known since H. C. Taylor and J. J. Russell published their tables in the *Journal of Applied Psychology* **23**, 565–578; 1939.

The validity coefficient is a multiplier in the numerator of the formula for the utility of a selection device. So if validity is zero, the test can have no utility. Owen's table is misleading as an example of low validity, as it corresponds to a (phi) correlation of 0.33 if 50% of applicants are suitable. This is as high a correlation as is typically claimed for good single tests. Our point is that in most cases test validities are seen to be not significantly different from zero when appropriate standards are applied.

S. BLINKHORN

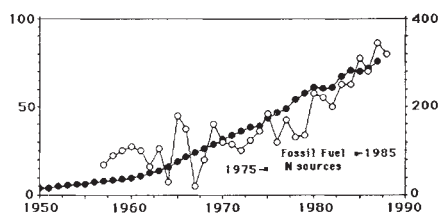
C. JOHNSON

Psychometric Research and
Development Ltd,
Brewmaster House, The Maltings,
St Albans, Herts AL1 3HT, UK

Fertilizer and climate change

SIR—Mayewski *et al.*^{1,2} reported a recent rise in nitrate concentration in carefully dated polar snow deposits (up to 1988), thereby extending our knowledge of the range and degree of global climate chemistry changes in remote environments, and discussed a link between these changes and both the formation of the Antarctic ozone hole and global warming. They suggested that the source of the precursor gaseous nitrogen ultimately deposited as nitrate came from fossil fuel sources. But, as I explain here, fertilizers could also be a source of recently deposited nitrate.

Global nitrogen fertilizer consumption



Nitrate concentration (O, ng per g) in snowfall at South Pole station T (Fig. 3 of ref. 1), world nitrogen fertilizer consumption (●, million mt) and an estimate of atmospheric nitrogen emissions from fossil fuels (■) in 1975 (ref. 5) and 1985 (based on fossil fuel consumption since 1975).

is increasing relative to the estimated release rate of nitrogen from fossil fuel sources (see figure). Global fertilizer consumption in 1975 (the date of the global atmospheric nitrogen budget used by Mayewski *et al.*) was about three times the estimated global emissions from fossil fuel sources. But from 1975 to 1985, global fertilizer nitrogen consumption increased about 62% (ref. 3) compared with an increase of about 60% for nitrate concentrations in snow, whereas global fossil fuel use increased only 21% (ref. 4). Global nitrogen fertilizer consumption in 1988 was therefore about 4.2 times larger than the probable sources of tropospheric nitrogen derived from fossil fuel emissions.

Much of the applied nitrogen fertilizer may return to the atmosphere within one year, especially in warmer climates. Estimates vary widely with soil type, but up to 80% of the applied nitrogen may be lost to the troposphere as gaseous nitrogen (such as NH_3 , N_2 and N_2O , refs 5–8). According to one rough estimate⁹, 36–50% of the fertilizer escapes to the atmosphere depending on whether fertilizer nitrogen goes into the plant or animal industrial food chain.

Extensive global consumption of nitrogen fertilizer began in the late 1950s, when nitrate concentration in ice from both the South Pole and south Greenland also began to increase rapidly (see figure). The correlation between nitrogen fertilizer

consumption and nitrate deposition at the South Pole station 'T' reported by Mayewski *et al.*¹ is very good (see figure). This correlation is consistent with the hypothesis, but does not prove it.

Thus, the nitrogen in fertilizers is a much larger potential source of troposphere gaseous nitrogen than fossil fuel sources. A significant amount of the fertilizer source is known to 'leak' into the troposphere, and the recent rise in nitrate concentration in dated Antarctic snow samples coincides with the rise in fertilizer consumption. If this hypothesis is correct, then a difficult additional dimension must be addressed to understand and respond to the complex problem of global climate change.

R. EUGENE TURNER

Department of Marine Sciences,
Louisiana State University,
Baton Rouge, Louisiana 70803, USA

1. Mayewski, P. A. & Legrand, M. R. *Nature* **346**, 258–260 (1990).
2. Mayewski, P. A. *et al.* *Nature* **346**, 554–556 (1990).
3. *Annual Statistical Yearbook* (United Nations, 1951–89).
4. *Energy Statistics Yearbook* (United Nations, 1986).
5. Ehhalt, D. H. & Drummond, J. W. in *The Tropospheric Cycle of NO_x* (eds Georgii, W. H. & Jaeschke, W.) 219–251 (Reidel, Boston, 1982).
6. Freney, J. R. & Simpson, J. R. (eds) *Dev. Plant Soil Sci.* **9**, (1983).
7. Rolston, D.E. in *Denitrification, Nitrification, and Atmospheric Nitrous Oxide* (ed. Delwiche, C. C.) 127–149 (Wiley, New York, 1981).
8. Bremner, J. M. & Blackmer, A. M. *Science* **199**, 295–296 (1978).
9. Delwiche, C. C. in *Denitrification, Nitrification, and Atmospheric Nitrous Oxide* (ed. Delwiche, C. C.) 1–15 (Wiley, New York, 1981).

Levitation of organic materials

SIR—We have succeeded in levitating at room temperature 'nonmagnetic' materials by means of a strong inhomogeneous static magnetic field. Such materials are in fact weakly diamagnetic and, when subjected to a magnetic field gradient, tend to be driven from regions of high field to those of lower field. When the resulting force is upwards and stronger than gravity, levitation occurs.

The critical criteria for levitation are the intrinsic magnetic property of the diamagnetic material (the specific magnetic susceptibility) and G , the gradient of the square of the magnetic field. For completeness, we also report B , the field at which the coils were driven to obtain such a gradient.

In the 5-cm cylindrical room-temperature bore of the hybrid magnet of the Service National des Champs Intenses (Grenoble), we have levitated various diamagnetic solids and liquids. Pure samples of bismuth and antimony were levitated with $G_{\text{bi}} = 729 \text{ T}^2 \text{ m}^{-1}$, $B_{\text{bi}} = 15.87 \text{ T}$ and $G_{\text{sb}} = 1,208 \text{ T}^2 \text{ m}^{-1}$, $B_{\text{sb}} = 18.75 \text{ T}$, respectively, values in very good agreement with calculations based on magnetic susceptibility data (R. C. Weast, *Handbook of Chemistry and Physics* 1972–73). Pieces of wood and plastic were levitated with $1,648 \text{ T}^2 \text{ m}^{-1} < G_s < 1,753 \text{ T}^2 \text{ m}^{-1}$, $21 \text{ T} <$

$B_s < 21.5 \text{ T}$ and $G_p = 1,923 \text{ T}^2 \text{ m}^{-1}$, $B_p = 22.28 \text{ T}$, respectively. Water, ethanol and acetone were levitated with $2,961 \text{ T}^2 \text{ m}^{-1} < G_w < 3,097 \text{ T}^2 \text{ m}^{-1}$, $26.5 < B_w < 27 \text{ T}$, $1,445 \text{ T}^2 \text{ m}^{-1} < G_e < 1,648 \text{ T}^2 \text{ m}^{-1}$, $20 \text{ T} < B_e < 21 \text{ T}$ and $1,862 \text{ T}^2 \text{ m}^{-1} < G_a < 2,086 \text{ T}^2 \text{ m}^{-1}$, $22 \text{ T} < B_a < 23 \text{ T}$, respectively. Values for the liquids were higher than expected and may result from wetting effects in the apparatus.

We have studied the levitation of graphite in a lower field magnet with a much larger bore ($G_g = 140 \text{ T}^2 \text{ m}^{-1}$ and $B_g = 5.25 \text{ T}$). We have confirmed that the levitation was very stable, without any contact with the magnet bore.

Our technique could be used to provide a contactless, microgravity environment for the elaboration of a wide range of materials. The case of organic materials is of great interest as they all have almost the same specific diamagnetic susceptibility, high enough to achieve levitation in superconducting magnets.

E. BEAUGNON

R. TOURNIER

CNRS, Centre de Recherche sur les
Très Basses Températures,
Laboratoire associé à l'université
Joseph Fourier,
25 Ave des Martyrs,
38042 Grenoble Cedex, France

Steady shine

SIR—Gry and Bonnet-Bidaud's claim¹ that the alleged ancient red colour of Sirius was due to the obscuring of Sirius by a globule is not corroborated by the historical record. There is no evidence for a fainter Sirius in the past. Ptolemy's brightness estimates are unreliable as the 14 other first-magnitude stars he listed range from stars as bright as Canopus ($V = -0.72$) to stars as faint as Denebola ($V = +2.14$) and Acamar ($V = +3.06$). In fact, both Manilius and Ptolemy identified Sirius as the brightest star in the night sky. Its unchanged brightness is further corroborated by ancient Babylonian and Hellenistic observations of its heliacal risings and settings. These phenomena depend on a parameter called the *arcus visionis*, which is correlated with the stellar brightness. The observations all result in an *arcus visionis* for Sirius comparable to its modern value.

The Chinese record by Simā Qiān (*Shiji*) is more correctly translated as: "To the east [of a previously discussed group of stars] there is a large star called Láng [= Sirius]; when Láng shakes its horns [scintillates?] and changes colour, there will be many robberies and thieves." According to the eighth-century commentator Zhāng Shōujié "Láng is a star, situated south-east of Shēn [= Orion]. Láng stands for a campaigning general and governs robberies and pillage. It is predicted that people turn against each other when Láng changes position. When the colour [of Láng] is yellow or white and bright, it signifies good fortune; if the horns [rays?] turn red, military disturbances will arise. When Venus, Jupiter or Mars guard [= become stationary near] Láng, the same applies." The *Shiji* lists many more prognostics based on colour and brightness changes of other bright stars, none of which is currently known to be intrinsically variable. Apparently, these variations were all due to atmospherical scintillation or extinction. In fact, Simā Qiān confirms that the undisturbed colour of Sirius resembles the colour of Venus.

Arguments have been made^{2–5} for an unchanged bluish-white colour for Sirius during the past few millennia. The occasional references to a reddish Sirius (Babylonian omen texts, Horace, Seneca, Ptolemy and so on) are explained more naturally as referring to observations of Sirius near the horizon for calendrical and astrological purposes⁶.

R. H. VAN GENT

Museum Boerhaave, Postbox 11280,
2301 EG Leiden,
The Netherlands

1. Gry, C. & Bonnet-Bidaud, J. M. *Nature* **347**, 625 (1990).
2. van Gent, R. H. *Nature* **312**, 302 (1984).
3. Tong, B. *Tang Nature* **319**, 532 (1986).
4. van Gent, R. H. *Nature* **325**, 87–89 (1987).
5. van Gent, R. H. *The Observatory* **109**, 23–24 (1989).
6. Rowan-Robinson, M. *Universe* p. 25 (Longman, 1990).

Stimulating patterns

Alan L. Mackay

Visions of Symmetry: Notebooks, Periodic Drawings, and Related Works of M. C. Escher. By Doris Schattschneider. *Freeman*: 1990. Pp. 354. \$39.95, £27.95.

EVERY student room used to have a print by Escher and a portrait of Che Guevara. Guevara was an obvious signal of revolt against authority, but Escher was also a revolt, perhaps against the fashionable art world and painters such as Jackson Pollack who, in retrospect, could charitably have been said to be exploring the regularities of chaos, by "flinging a pot of paint in the public's face". Students, and others like Tom Keating, enjoy mocking the pretentious art critics and seize on an artist who explores serious problems and who has a masterly command of nontrivial techniques.

Maurits Escher (1898–1972) has become a highly regarded artist, in spite of the orthodox art world. Escher appealed to those who, in a previsual age, would have been enthralled by H. G. Wells, Jules Verne and other science fiction authors scorned by the literary pundits who had little understanding of science and technology and their possibilities. Douglas Hofstadter, in his eccentric masterpiece *Goedel, Escher, Bach; An Eternal Golden Braid* (Basic Books, 1979), coupled Escher with Goedel and Bach as a source of stimulating patterns.

Seurat explored the representation of colour by points (now called pixels) stimulated by scientific theories; Brunelleschi developed projection. Escher explored the laws of symmetry and projections and especially the contradictions of perception consequent on the representation of three dimensions with two. He invented transformations of one tiling to another.

Freedom is the recognition of necessity and we might ask what freedom Escher, the Netherlands print-maker had? Curiously, very few of his prints are concerned with his native landscape and early in his career he escaped to the Mediterranean. He enjoyed the bright light of Calabria and meticulously noted the multicoloured Islamic tilings of the Alhambra. It is natural for a print-maker to work backwards and forwards between the block and its mirror image. Perhaps pre-war life in Holland was black and white, Protestant or Catholic, with no third choice. Quite early Escher discovered the regular tilings of flat two-

dimensional space with permutations of colours and found them "the richest source of inspiration I have ever struck." It is a surprise that Delft does not manufacture quantities of Escher tiles and that only a few special commissions were executed in real tiles.

Escher made extensive investigations of the restrictions of plane symmetry, and by 1926 had worked out for himself much about the plane groups. He was introduced in 1937 by his brother to George Pólya's paper on the 17 crystallographic



Interlocking fish drawn by Escher in March 1938.

plane groups of symmetry and to Heinrich Heesch's works on black and white symmetry, but he preferred the job of working things out for himself to mastering the esoteric mathematical literature. During the war, Escher lived in his small home town of Baarn, and life in German-occupied Holland was very grim. Escher's art teacher, S. J. de Mesquita, and his family, were taken away and exterminated. The last winter of the war brought dreadful famine. Escher kept his head down and explored the rigorous restrictions of the 17 plane groups, the 46 two-dimensional black and white groups, and the plane coloured groups, creating his own world. Curiously, Heinrich Heesch, who developed these black and white groups, worked for the Todt Organisation during the war, showing how these symmetrical plane packings could be used to stamp components more efficiently from steel strip.

After the war, Escher's work quickly came to the notice of Roger Penrose, P.

Terpstra, J. D. H. Donnay and Gabrielle Donnay, Norman de Bruijn, Caroline MacGillavry and other mathematicians and crystallographers. A major exhibition at the Stedelijk Museum was organized by de Bruijn for the 1954 International Congress of Mathematicians and from that time Escher became a producer of icons for the scientific world. He was a guest of honour at the Cambridge Congress of the International Union of Crystallography in 1960. It pleased him greatly to be a bridge between the two cultures.

From the first, crystallographers have been aware of the decorative possibilities of planar and solid symmetry. This formed a big feature of the Festival of Britain in 1951. The Festival Pattern Group's principal consultant was Helen Megaw and the enterprise had been stimulated by an address in May 1949 given by Kathleen Lonsdale to the Society of Industrial Artists. Lawrence Bragg wrote (11 May 1951): "When in 1922 I worked out the first crystal of any complexity that had been analysed, aragonite, I remember well how excited my wife was with the pattern I showed her as a motif for a piece of embroidery". Escher's prints of interlocking animals were even brought to the notice of the Otovalo Indians of Peru and Ecuador, who had been making periodic patterns from pre-Columbian times and who added them very naturally to their repertoire of textiles.

The notebooks, now published here, with extensive commentary, contain several additions to the Escher corpus, which, like the Köchel numbers for Mozart's works, now carry Bool numbers, reaching 448. They document meticulously Escher's development.

Why the fascination of the regular polyhedra? From Plato onwards many people have been captivated by them and there is a current boom in symmetry. Even Mao Tse-Tong asked (of Tsungdao Lee, who had used Escher's most ingenious "Horsemen" on the cover of his colleague C. N. Yang's book *Elementary Particles* (Princeton University Press, 1962) "Tell me, why should symmetry be of importance?" In fact the concept of a group of operations is one of the most fundamental abstract discoveries. Escher found many of the symmetry groups for himself, perhaps finding some new groups of colour symmetry, and worked out many surprising consequences. Unfortunately he was not able to do more than make a beginning on the properties of spaces with other metrics and the rise of computer graphics has somewhat overshadowed

more manual skills. The ability to cut blocks which print accurately together has now been devalued by the advent of the machine.

Escher has stimulated some pure mathematics and Huson, Delgado and Dress (Bielefeld) have developed a symbolism which enables patterns of two kinds of tiles to be classified and generated. Their programme includes a kind of Escher machine which shows all the consequences of adjusting the perimeter between two tiles.

Escher's last print (of serpents entwined in interlocked rings) explored the consequences of an average coordination number greater than six. It is also a technical masterpiece of block cutting and printing. Escher was being prompted, by Coxeter, Penrose and others, to look at more general geometries. Penrose's

tilings which force nonperiodicity came too late, as did the flood of curved surfaces, such as the periodic minimal surfaces, now in spate, thanks to the advances in computer graphics.

Even crystallographers have much to learn. This book contains very many colour reproductions of the periodic drawings and analyses the 1941–42 notebooks which show Escher's development in this direction. Taking Doris Schattschneider's beautiful volume with earlier books, especially that by Bruno Ernst, documentation of Escher's life, intellectual development and corpus must now be almost complete. It makes a distinctive island in art history. □

Alan L. Mackay is in the Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

The end of innocence

John W. Birks

Global Air Pollution: Problems for the 1990s. By Howard Bridgman. Pinter: 1990. Pp. 261. Hbk £30, \$49; pbk £12.95, \$25.

The Changing Atmosphere: A Global Challenge. By John Firor. Yale University Press: 1990. Pp. 145. \$19.95, £12.95.

CHANGES taking place in the atmosphere today as a result of fossil fuel combustion, chemical manufacturing, agricultural practices and deforestation have already had well documented detrimental effects on many of the ecosystems we depend on, but the effects experienced so far are small in comparison to those projected for the future. These two timely books treat acid rain, global warming, stratospheric ozone depletion, photochemical smog, and other global environmental problems mediated by the atmosphere in highly complementary ways.

Global Air Pollution is written in the style of a textbook suitable for advanced courses in applied climatology, environmental science or geography. It does not have sufficient chemical detail for a course in atmospheric chemistry, but would serve as an excellent supplement to other texts. It is replete with figures, tables and an extensive bibliography. The book treats all aspects of air quality problems — tropospheric and stratospheric, gas phase and condensed phase, health-related and climate-related — in an even-handed way. Earlier texts in this field have tended to emphasize predominately air pollution (troposphere, smog, aerosols) or aeronomy (stratosphere and above, ozone depletion, gas phase only), representative of two communities of scientists that

interact too infrequently. Topics of biogeochemical cycling, atmosphere-biosphere interactions and global warming, all included in this book, have either not been treated or have been given only cursory attention in earlier texts. I was particularly pleased to find a chapter devoted to nuclear winter and was anxious to see how a scientist not involved in the debate would summarize the findings. I found the treatment of the subject to be up to date, fair, and to come out about where I do in my assessment of the most probable climatic impacts and associated uncertainties of a major nuclear war.

The Changing Atmosphere is written for the nonspecialist, and its 125 pages of text can be read in one or two sittings. Even for the atmospheric scientist, however, this is a very interesting book, partly because it provides novel ways of describing concepts such as steady states, numerical modelling, and radiochemical dating to the nonscientist, but more importantly, because it discusses possible solutions to our dilemma. Many atmospheric scientists, including myself, are much better at describing the problems of the atmosphere in detail than at proposing reasonable solutions. But John Firor has been thinking about environmental problems in the broader context of history, energy, economics, politics, conservation, population control and even the subjugation of women. He asserts: "Acid rain, ozone depletion, and climate heating — the three most highly publicized aspects of the rapidly changing composition of the atmosphere — are more than just related. A single, larger problem underlies them. They are all three the result of the impact of human activities on the earth, which now equals and even exceeds the influence of large-scale natural forces." Firor points out that our species is not different from any other species in its modification of the environment: trees drop leaves, foxes dig burrows, beavers cut trees, algae

turn lakes into bogs. But these species have been held in check by disease, accidents, predators and a limited food supply. Humans differ in that they have been able to shake off some of these constraints, but paradoxically, the waste products of our success threaten the destruction of other species and ancient ecosystems on which our health and progress depend.

Firor describes two possible paths that humanity may follow. We may continue along the path of subjugation of the Earth, as decreed by God in the book of Genesis, in which case we must be "ready to embark on global management of the soil, the air, the oceans, and those species we decide should survive." But as he aptly points out: "For a people who are not yet able to describe fully the workings of a tree or the chemistry of an ocean or to put available food in the hands of those who need it, it would be perilous to undertake the design and maintenance of a complete life-support system for everyone." The alternative path of development he recommends is one in which we lessen our impact on the atmosphere enough to allow both ecosystems and human civilization to adapt to the changes. He warns that we cannot rely on a technological solution; there is no cheap source of energy just around the corner. Instead, he provides a recipe of increased energy efficiency for both industrialized and developing countries, protection of forests, and stabilization and possibly reduction in the human population. Firor does not have all the answers, but his analysis identifies the roots of the problems and provides us with a starting point.

Bridgman reaches some of the same conclusions in the final chapter of *Global Air Pollution*, which also deals with strategies for avoiding irreversible climate change. Needs for control over the world population, energy conservation and efficiency, sustainable development and lessening the gap between the rich and the poor are emphasized. This is the first text on atmospheric chemistry I have seen that considers the links between atmospheric chemistry and topics such as population, world hunger, gross national products, the World Bank, and genetic diversity of rain forests. Bravo! □

John W. Birks is in the Department of Chemistry and Biochemistry and Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado 80309-0216, USA.

New in paperback

■ *Vygotsky's Psychology: A Biography of ideas* by Alex Kosulin. Published by Harvester Wheatsheaf at £12.95.

■ *First the Seed: The Political Economy of Plant Biotechnology, 1492–2000* by Jack Ralph Kloppenburg Jr. Published by Cambridge University Press at £12.95, \$14.95.

Visual effects

C. J. S. Clarke

Beyond the Third Dimension. By Thomas Banchoff. Freeman: 1990. Pp. 210. £16.95, \$32.95.

At first glance, *Beyond the Third Dimension* might seem to be a trendy coffee-table book, striving for attention through its computer graphics effects. But in reality Thomas Banchoff has succeeded in producing one of the best popular books on geometry that I have seen. The approach is deceptively simple, based on numerous illustrations, both computer-generated and conventional, that are carefully linked with the text in a wide-margined format. The visual effects are indeed strikingly beautiful, but strictly kept at the service of explanation. Co-ordinate geometry is postponed to a late chapter, and yet most of the main themes of modern geometry are touched on, giving a well rounded and concise account of geometry from the viewpoint of the idea of dimension.

A particular strength is the attention given to applications. All the main points are illustrated by concrete examples, such as the manipulation of circular pools of light on a stage as an illustration of the geometry of a space of circles. In addition, there are practical applications, such as studies of human movement made to understand lower back pain ("the geometry of rehabilitation therapy") and an extended case study of the use of multi-dimensional ideas in palaeoecology: understanding the way in which climate has changed in the US midwest by measuring the abundance of different sorts of pollen grains in ancient sediments.

The key theme, providing a unifying thread of examples and illustrations, is that of the regular solids in three and higher dimensions; but the coverage includes fractals (fractional dimensions), manifolds, exotic differentiable structures, Morse theory and caustics. It was also refreshing to see the subject placed in its historical context, including less well known figures such as Friedrich Froebel, the inventor of the kindergarten, along with Egyptian texts, Gauss and Kant.

The depiction of structures and datasets in higher dimensions is currently an active area of computer science research, on which an interesting light is shed by the approach of this book. It is notoriously difficult to give a usable encoding of four-dimensional, and higher, information by adding colour, texture and so on to computer images, and this is very rarely attempted here. Instead, all the images are explicitly of three-dimensional objects, but great care is taken to ensure that the viewer understands the formal

relationship between the three-dimensional picture and the four-dimensional structure from which it derives. The real key to higher-dimensional understanding, however, is the use of movement as a means of combining and relating many different three-dimensional views of the underlying structure. This is stressed and described by the author, but cannot be conveyed in a static book. Perhaps, as hyper-media presentations become more developed, we can look to a computerized sequel to this book in a few years time which will give us this additional dynamic ingredient. Meanwhile, all readers, whether lay or professional mathematician, can expect to learn something new from this book. □

C. J. S. Clarke is in the Faculty of Mathematical Studies, University of Southampton, Southampton SO9 5NH, UK.

Three's a crowd

John Galloway

Three, Four and More: A Study of Triplet and Higher-Order Births. Edited by Beverley J. Botting, Alison J. McFarlane and Frances V. Price. HMSO: 1990. Pp. 246. Pbk £12.95.

IN Britain today one in 45 children is a twin, a triplet or a higher multiple (a rather clumsy term but it is difficult to think of a better one). Numbers are rising as a result of treatment for infertility and the increasing survival rate of premature babies, a group to which many multiples belong.

The difficulty small minorities face is often simply a lack of knowledge about them. Individual doctors, health visitors, teachers and parents may have little or no first-hand experience of multiples and may not know how to go about drawing on the experience of others. A lack of real information about multiples all too readily leads to society merely being curious about them rather than seeing them as individuals with a set of unusual problems — and opportunities — arising from the special circumstances of their birth.

Three, Four and More is the result of research into triplets and higher multiples and the problems they, their families and professional helpers face. The statistics which have been assembled are being used to mount a campaign to help them. No comparable book has ever been written for twins and indeed until 1983 there were no books at all for the general public published in Britain; although the Twins and Multiple Births Association (TAMBA) had published regular magazines and leaflets on the care of twins from their formation in 1978. TAMBA's aim has been to pool the knowledge and experience of

those who really know about multiples — their parents — and to use that information in turn to help multiples and their families. This was the first such organization in Europe although they already existed in other countries, Australia and North America for instance.

Multiples are really rather rare but just how rare are they? In Britain, about one birth in 80–100 produces twins. So roughly one child in 40–50 is a twin and a small primary school will expect to have a couple of sets of twins at any time.

The frequency of naturally occurring higher-order births seems to bear some

The Caen quads — such births used to be one in a million but all that has changed with modern infertility treatments.

sort of regular relationship to that of twins. It used to be suggested that the relative frequency of triplets was the square of that of twins, that of quadruplets the cube. So one birth in 10,000 would be a triplet, and one in a million a quadruplet birth. With fertility drugs, *in vitro* fertilization (IVF) and GIFT treatments these statistics no longer apply, and the rapid increase in higher multiple births emphasizes the importance of *Three, Four and More*. The study on which the book is based ran for five years, starting in 1980, when hormones were used for treating infertility but IVF had not yet had an impact. Unlike most *twin* studies, this one concentrates on the needs of parents, and of professionals. Even a neonatal paediatrician, who specializes in the hospital care of preterm babies, may not see a triplet or higher-order multiple birth every year.

The report is not intended to be a definitive work — the authors hope it will spark off more research, for example on the value of compulsory bed rest during pregnancy. Mothers of triplets spent an average of 29.5 days in hospital before delivery, mothers of quads 54.4 days, compared to 9.5 days for twins and 1.4 days for singletons. But there is no evidence to suggest that routine hospital bed rest leads to a better outcome, and indeed

a recent report in the *Lancet* has called for an end to routine antenatal admissions for women with twin pregnancies if the pregnancy is 'uncomplicated'. This follows an Australian study which showed that mothers who were randomly allocated outpatient care from the 26th week were less likely to have high blood pressure, discomfort and premature labour, compared with mothers randomly made to stay in hospital, especially if they had other children at home. This is presumably the sort of advice which underfunded hospitals will be happy to follow.

Of course, doctors have been advising bed rest in the hope of preventing a premature delivery. This is one of the main hazards of a higher multiple birth, resulting in problems for everyone concerned. For the babies there is an increased risk of cerebral palsy, with children who weigh less than 1,500 grams (3 lb 5 oz) at birth being most at risk, (a rate of 68 per 1,000 compared with 2 per 1,000 in the general population). In fact all triplets and higher multiples are more at risk of cerebral palsy (17.4 per 1,000 children), whatever their birth weight. Low-birth-weight children are also more likely to be readmitted to hospital for other reasons.

For the parents, having three or more premature babies in special care units causes stress; and if they are split up between various hospitals as they often are, the logistics of visiting them may be impossible, especially because the mother may be far from well. Over a quarter of mothers of triplets and over a third of mothers of quads have medical problems after the birth which include haemorrhage, anaemia, infections and high blood pressure.

For hospitals working on a tight budget the number of intensive-care cots has to be kept down. In fact over the country the number has actually been reduced. And it is extremely unlikely that three will be empty in the same hospital at the time when triplets are born. The number of sets of triplets born each year has more than doubled with the emergence of private IVF clinics. Conceived profitably by private medicine, the problems of neonatal care fall on a public health service that is short of money. The report estimates that at 1988 prices the National Health Service cost of hospital care is £5,000 for a set of twins, £12,000 for triplets and £25,000 for quads or more.

Three, Four or More makes compelling reading with its combination of statistics and quotes from parents, the cost to whom in money, energy and emotion is enormous. But relationships between multiples are among the strongest and most enduring of any human relationships. □

John Galloway is at the Cancer Research Campaign, 2 Carlton House Terrace, London SW1Y 5AR, UK.

Ten years on

B. Hogan

Transgenic Animals: Edited by N. L. First and F. P. Heseltine. *Butterworth-Heinemann*: 1991. Pp. 358. £60, \$75.

A DECADE has now passed since the first transgenic mice were generated by the microinjection of cloned DNA into the pronuclei of fertilized eggs. The production of transgenic mice is now almost a routine procedure, and microinjection has also been used to introduce foreign DNA into rats, pigs, sheep and cows. More recently, it has become clear that the targeting of mutations to specific genes by homologous recombination in murine embryonic stem (ES) cells, although technically much more difficult than microinjection, also has the potential to become a standard experimental tool. At a time when many scientists are dreaming of future possibilities opened up by the intrepid pioneers of ES cell manipulation, it is useful to step back a little and reflect on what has been achieved during the decade of research with transgenic animals. What hopes have been fulfilled and what unexpected insights gained? Equally important, what problems remain unsolved and what has proved to be more difficult than expected? How will ES cell technology answer questions that cannot be addressed using microinjection? A review volume, covering these and other topics would be very useful, and *Transgenic Animals* goes some way towards providing a broad overview for people in medical and agricultural research interested in learning just what transgenic technology has to offer them.

Unfortunately, though, the contributions to this book are rather uneven and the potential reader must be prepared to pick and choose from a mixed offering. The problem arises from the fact that *Transgenic Animals* is the proceedings of a workshop convened in Bethesda in 1988 by the National Institute of Child Health and Human Development. The staff of NICHD have been staunch supporters of research using transgenic and ES cell technology, and the goal of the meeting was to bring the achievements and potential benefits of transgenic research to the attention of a wide audience, and to place the research within the context of other work on gene regulation, gene therapy and models of human disease. Inevitably, some of the speakers presented only current research results which are now rather dated. It would have been better if these papers could have been omitted in favour of expanding others in which experimental data are presented clearly within the context of general principles.

One of the major achievements of

transgenic technology has been the ability to define DNA sequences controlling complex temporal and tissue-specific patterns of gene expression. In many cases, the delineation of positive and negative regulatory elements would have been impossible without transgenic technology. Perhaps the most celebrated example is the discovery of the locus control region (LCR) far upstream of the human β -globin gene cluster. This LCR is required for high levels of expression of the different β -globin genes in a developmentally correct sequence, and is thought to influence the structure and accessibility of chromatin over a considerable distance. Other examples of tissue-specific gene expression are reported in this volume, including studies with genes for α -fetoprotein, histocompatibility antigens, and milk and muscle proteins, thus adequately covering a range of regulatory strategies.

Another theme of *Transgenic Animals* is the application of transgenic technology to the production of models for human disease. The goal has been to generate animals which can provide real insight into complex disease processes and act as a test bed for therapeutic intervention, rather than just pathetic examples of pathology and deformity. In some cases success has been achieved, for example in producing models for multistage progression in cancer or for testing vaccines. In other cases mouse models have been disappointing, either because they have produced no real insight or because the disease phenotype was not obtained. For example, homozygous HPRT⁻ mice show none of the symptoms of Lesch-Nyhan disease because of a fundamental difference in purine metabolism between mice and humans. Overcoming such problems requires the ability to introduce genes into a range of experimental animals, a topic well covered in this book. The recent observation by R. E. Hammer and colleagues that transgenic rats carrying the human HLA-B27 gene display the symptoms of the complex inflammatory disease ankylosing spondylitis, whereas transgenic mice do not, testifies dramatically to the importance of a multi-organism approach. The pluses and minuses of introducing transgenes into farm animals for agricultural improvement are also adequately addressed, and the book concludes with a chapter on the ethics of patenting transgenic animals. In summary, this is not a 'how to' book. Nor does it provide a comprehensive account of all achievements using transgenic animals. Rather the articles are a reflection of research projects as they really are; some good, some indifferent, and some so good they make it all worthwhile. □

B. Hogan is in the Department of Cell Biology, Vanderbilt School of Medicine, Nashville, Tennessee 37232, USA.

The conformation of membranes

Reinhard Lipowsky

Membranes composed of amphiphilic molecules are highly flexible surfaces that determine the architecture of biological systems and provide a basic structural element for complex fluids such as microemulsions. Physical theories have been developed to describe many aspects of their conformational behaviour, such as the preferred shapes and shape transformations of closed vesicles, and the shape fluctuations, random-surface configurations, and adhesion and unbinding of interacting membranes. Understanding of these phenomena has been much improved through fruitful interactions between theory and experiment.

THE membranes that I shall consider here are very thin and highly flexible sheets of amphiphilic molecules—structures that are a ubiquitous component of biological systems. On length scales large compared with the size of the molecules, these membranes can be regarded as two-dimensional surfaces embedded in three-dimensional space. The behaviour of these surfaces has recently attracted interest from several communities.

Biology, biophysics and biochemistry. Membranes represent the main structural component for the complex architecture of biological systems. The human brain, for example, consists of a complex network of membranes with a total surface area of 10^3 – 10^4 m². Biophysicists have constructed a number of model membranes that are expected to capture some of the essential features of their biological counterparts. The simplest models of this kind are provided by lipid bilayers which assemble spontaneously from lipid molecules dissolved in water.

Physical chemistry and chemical engineering. Surfactant molecules in mixtures of immiscible fluids often assemble into membranes which then form a variety of different structures. These supramolecular structures can include: a liquid crystal such as a smectic phase formed by an ordered stack of membranes; a periodic crystal in which the membranes divide space into two interpenetrating labyrinths and thus form a bicontinuous structure; or new, unusual phases such as the so-called sponge phase, which represents a bicontinuous but disordered state of matter.

Theoretical physics and mathematics. To understand the conformational behaviour of membranes, one has to introduce a number of novel theoretical concepts such as bending elasticity and curvature, scale invariance arising from fluctuations on many length scales, and renormalization of membrane interactions. I will try to explain these theoretical concepts and their relation to experiments. One intriguing aspect of this topic is the need to combine mathematical techniques such as differential geometry (see Box A) and renormalization group theory (see Box B) with the behaviour of real systems as observed under the microscope.

Mesoscopic shape of membranes

In aqueous solution, lipid or surfactant bilayers typically form closed surfaces or vesicles. Multilamellar vesicles consisting of several phospholipid bilayers were first observed by electron microscopy¹. Several methods are now available through which one can obtain large unilamellar vesicles consisting of a single lipid bilayer². Such vesicles have a linear size of the order of 1–10 μ m and can be studied directly by optical microscopy. These observations have shown that lipid vesicles exhibit a large variety of different shapes (Fig. 1*a, b*); in particular, they can exhibit the non-spherical, biconcave shape typical of red blood cells^{3,4}.

Bending elasticity and formation of vesicles. It is well known that liquid droplets usually have a spherical shape which is governed by interfacial tension. Non-spherical shapes cannot be determined by such a tension. It is now generally believed that the shape of vesicles is determined primarily by bending

elasticity and curvature^{5–7}. This idea leads to the theoretical model described in Box A. In this model, which I will call the curvature model, the energy of a given shape depends on the mean curvature and on the gaussian curvature of the membrane⁶ (see Box A). In its simplest version, this model involves only two parameters: the bending rigidity, κ , and the gaussian curvature modulus, κ_G . For phospholipid bilayers, $\kappa \approx 10^{-19}$ J (refs 8, 9).

One nice feature of the curvature model is that it provides a very simple explanation for the fact that bilayers tend to form vesicles. Consider a membrane segment of linear size L . If this segment is planar, it has no bending energy but its boundary has an edge energy, σ_E , per unit length arising from the partial contact between the water and the hydrocarbon chains of the lipid. The total edge energy of the segment is thus proportional to L . On the other hand, if the segment forms a sphere, it has no edge energy but the bending energy is non-zero, and is given

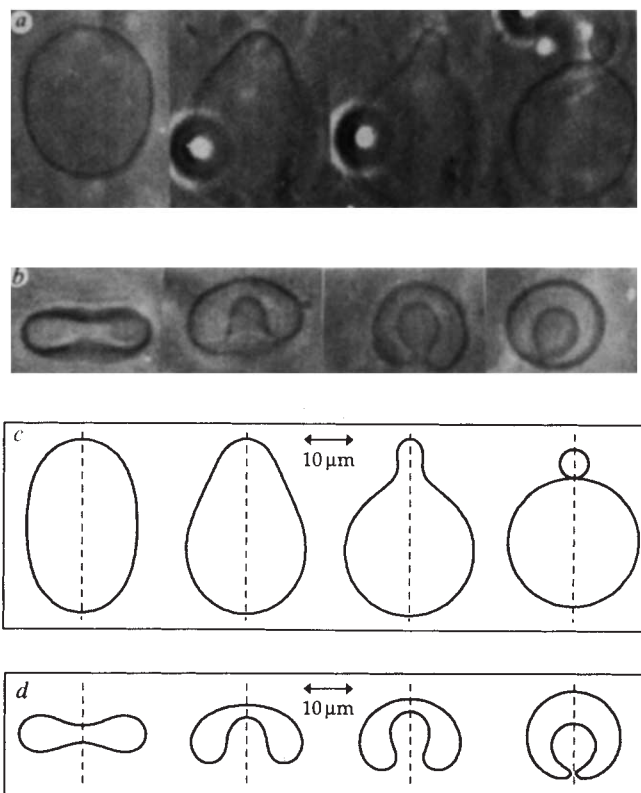


FIG. 1 Shape transformations of free vesicles induced by a change in temperature⁴. *a, c*, Expulsion of a small vesicle from a larger one (budding). *b, d*, Inverse budding ('endocytosis') via the transformation from a discocyte to a stomatocyte. The shapes are axisymmetric with respect to the broken line.

by $4\pi(2\kappa + \kappa_G)$, which does not depend on the linear size L of the membrane (see Box A). Therefore, for large L , the membrane can always lower its energy by forming a closed surface. For phospholipid bilayers, the edge energy $\sigma_E \approx 10^{-20} - 10^{-19} \text{ J nm}^{-1}$ (refs 10, 11). This mechanism of vesicle formation via the hydrophobic effect is quite general, and operates in biological systems which usually contain a large number of such structures.

Shape transformations of free vesicles. Consider again the vesicle shapes shown in Fig. 1a and b. These micrographs show a single lipid vesicle, which undergoes shape transformations induced by a change in temperature.

In thermal equilibrium, the vesicle should attain the shape that corresponds to the minimum bending energy. If one assumes rotational symmetry, these minimal shapes can be calculated from the curvature model¹²⁻¹⁴. It has been found recently that, if one assumes that the two monolayers of the bilayer exhibit a small asymmetry^{4,14}, the experimentally observed shapes and shape transformations shown in Fig. 1a and b can be explained by such minimal shapes (Fig. 1c, d).

Adhesion of vesicles and vesicle fusion. Now consider a vesicle that interacts with another membrane or surface. If this interaction is attractive, it can lead to a bound state of the vesicle. Bound states have been investigated experimentally by micropipette techniques¹⁵, which allow one to study, for example, the adhesion of two vesicles (Fig. 2).

Box A. DIFFERENTIAL GEOMETRY AND THE CURVATURE MODEL

Consider a two-dimensional surface embedded in three-dimensional space. At each point on the surface, one can define two principal curvatures, C_1 and C_2 , which determine the mean curvature $C_1 + C_2$ and the gaussian curvature $C_1 C_2$. The curvature model is then defined by the configurational energy or 'effective Hamiltonian'⁶

$$\mathcal{H} = \oint dA \left\{ -\kappa C_0 (C_1 + C_2) + \frac{1}{2} \kappa (C_1 + C_2)^2 + \kappa_G C_1 C_2 \right\} \quad (\text{A.1})$$

where the integral extends over the whole membrane surface (which could consist of several disconnected components) and dA represents the intrinsic area element. The parameters κ , C_0 and κ_G are the bending rigidity, spontaneous curvature and gaussian curvature modulus, respectively. For a symmetric membrane (both sides identical), $C_0 = 0$.

The shape of a vesicle with surface area A and enclosed volume V is determined by minimization of $\mathcal{H} + PV + \Sigma A$, where P and Σ denote the difference between the outside and the inside pressure and the lateral tension, respectively. For phospholipid bilayers, the exchange of lipid molecules with the surrounding solution is very slow and the area A is essentially constant on experimentally relevant timescales. In addition, one may control the vesicle volume V , for example by varying the concentration of a solute that cannot permeate the bilayer. The pressure P and the tension Σ then play the role of Lagrange multipliers. This approach has been used to calculate the shapes shown in Fig. 1 and Fig. 3.

For a spherical vesicle of radius R , $|C_1| = |C_2| = R$. When inserted into (A.1) with $C_0 = 0$, one obtains $\mathcal{H} = 4\pi(2\kappa + \kappa_G)$. In fact, for a closed surface, the third term in (A.1) always leads to a constant that does not depend on the size of the surface but only on its topology or, more precisely, on its Euler characteristic⁹⁶, χ . For a surface without edges, the integral over the gaussian curvature is given by

$$\oint dA C_1 C_2 = 2\pi\chi \quad (\text{A.2})$$

as follows from the Gauss-Bonnet theorem of differential geometry⁹⁶.

Now consider the minimal surface shown in Fig. 4b, for which the mean curvature, $C_1 + C_2$, is zero everywhere. This surface is topologically equivalent to a cubic lattice of spheres connected in all three spatial directions by pieces of cylinders; it has Euler characteristic $\chi = -4$ per lattice site. Thus, for $C_0 = 0$, the bending energy of the minimal surface in Fig. 4b is given by $\mathcal{H} = -8\pi\kappa_G$ per lattice site. The same energy applies, for $C_0 \neq 0$, to a surface with constant $C_1 + C_2 = C_0$. The latter surface represents a minimum of $\mathcal{H} + PV + \Sigma A$; it thus represents a model for the prolamellar body shown in Fig. 4a.

The shape of a bound vesicle is determined by the interplay between adhesion and bending energies. This interplay can be studied theoretically by starting from a simple generalization of the curvature model in which the membrane experiences a contact potential arising from the attractive surface^{16,17}. If one assumes that the vesicle does not change its topology, theory predicts a large variety of different bound states (Fig. 3). As shown in this figure, the vesicle can undergo shape transformations between free and bound states as well as transformations between different bound states.

Adhesion can also lead to topological changes such as vesicle rupture and vesicle fusion. In the limit of strong adhesion, a bound vesicle with constant volume attains the shape of a spherical cap. In this limit, the adhesion energy, E_{ad} , is related to the lateral tension Σ by the Young-Dupre equation: $E_{ad} = \Sigma(1 + \cos \psi_{eff})$, where ψ_{eff} denotes an effective contact angle^{16,17}. If the tension Σ exceeds a certain threshold, it will disrupt the membrane. As a result, the closed vesicle will be transformed into a disk-like membrane segment with a free edge. A segment of linear size L has an adhesion energy proportional to L^2 , which overcomes the edge energy (proportional to L) for sufficiently large L .

If two vesicles collide, they may adhere and eventually fuse into a larger one¹⁸. According to the curvature model, the energy gained by the fusion of two free vesicles is $\Delta E = 4\pi(2\kappa + \kappa_G)$ ¹⁹. Thus, fusion of two free vesicles is energetically favourable if $\kappa_G > -2\kappa$. In principle, the fusion process also involves a loss of translational entropy which leads to a free-energy increase of the order of $kT \ln(R/L)$ where R and L are the linear sizes

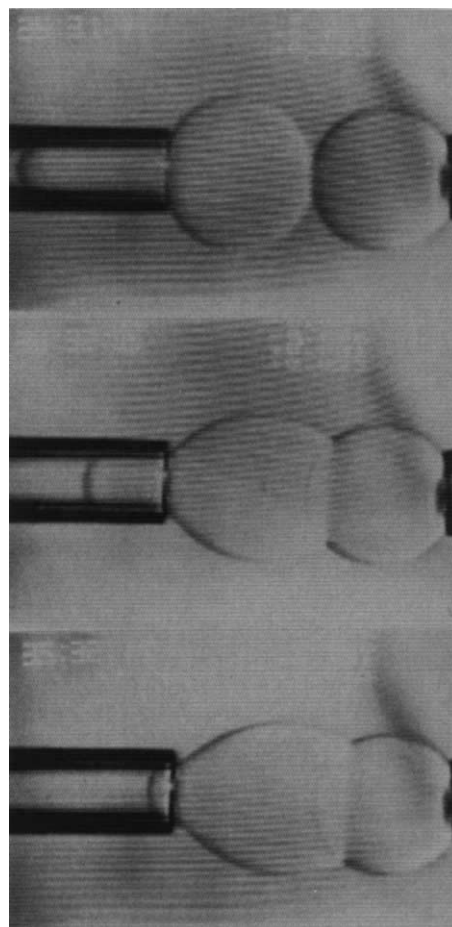


FIG. 2 Adhesion of two lipid vesicles which are brought into contact by two micropipettes. The vesicle radii are $\approx 10 \mu\text{m}$ (courtesy of E. Evans).

of the container and of the vesicle respectively (k is the Boltzmann constant). For lipid bilayers with $\kappa \approx 10\text{--}20\text{ kT}$, this entropic contribution is usually small relative to the bending energies.

On the other hand, for two bound vesicles of linear size L , the energy gained by fusion is proportional to L for large L (ref. 17). Therefore, two vesicles can fuse in their bound state even if they cannot fuse in their free state. Such an adhesion-induced fusion process has been observed for vesicles attracted by a gold surface²⁰ and for lecithin membranes at the air-water interface²¹.

Ordered and disordered assemblies of bilayers. So far, I have discussed the behaviour of one or two vesicles in plenty of water. For a larger concentration of amphiphilic molecules, membranes can assemble into a variety of different structures. First, the membranes may spontaneously form an ensemble of unilamellar vesicles, as has been observed for bilayers of several kinds of surfactant^{22–25}. Lipid bilayers, on the other hand, often form layered structures such as multilamellar vesicles, myelin cylinders or oriented stacks, which are stabilized by interactions between the membranes.

Another possible structure is a triply periodic surface which divides the whole three-dimensional space into two interwoven labyrinths or subvolumes and thus gives rise to a bicontinuous phase^{26–30}. Such crystalline states of bilayers have been reported for a variety of amphiphilic molecules^{27,29,30} and are present in plant cells²⁶ (Fig. 4a). In some cases, the partitioning bilayer forms a 'minimal surface', which is characterized by zero mean curvature. An example with cubic symmetry is shown in Fig. 4b. Minimal surfaces can exhibit many different symmetries, and there has been recent progress in classifying the corresponding space groups³¹. Triply periodic surfaces with constant but non-zero mean curvature have also been constructed^{32,33}.

The ordered structures shown in Fig. 4a and b are perturbed by thermally excited shape fluctuations of the bilayers. These fluctuations can pinch off some of the narrow necks and can introduce other types of defects, causing the crystals to melt into a sponge (Fig. 4c). A sponge phase, which is still bicontinuous but exhibits no long-range order, seems to have been observed for surfactant bilayers^{34–36}. This interpretation is still somewhat controversial: an alternative model in terms of disk-like membrane segments has also been proposed (C. A. Miller *et al.*, preprint, Univ. of Bayreuth).

Shape fluctuations of membranes

In general, low-dimensional objects such as interfaces, membranes and polymers undergo thermally excited shape fluctuations to increase their configurational entropy³⁷. A membrane is usually more flexible than an interface but less flexible than a polymer. It has been realized recently that the character of its shape fluctuations depends on the internal state of the membrane, which can be fluid, polymerized or hexatic (see below).

The shape fluctuations of a lipid vesicle can be observed directly under a light microscope. In this way, one can probe fluctuations with wavelengths of the order of the vesicle radius. It is important to realize, however, that these fluctuations involve a large range of different length scales, from molecular dimensions up to the membrane size, most of which cannot be resolved by the microscope.

In the following, I will assume that there is essentially no reservoir of amphiphilic molecules within the aqueous solution—in other words, that these molecules have all assembled into bilayers. This should apply to lipids and double-chained surfactants that have a very low critical micelle concentration. One is thus led to study the shape fluctuations of random surfaces that contain a fixed number of molecules.

Fluid membranes. Biological membranes are typically fluid³⁸, which means that the molecules can diffuse rapidly within the membrane. This fluidity was first recognized in experimental studies of lipid bilayers³⁹. At sufficiently high temperatures,

these bilayers exhibit a fluid phase with a diffusion constant $D \approx 10^{-7}\text{--}10^{-8}\text{ cm}^2\text{ s}^{-1}$ (ref. 40).

The molecules within a fluid membrane possess no reference lattice. In the absence of lateral tension, the shape of the membrane is then governed by its bending rigidity, κ , and the shape fluctuations represent bending modes. These modes lead to a certain roughness: at temperature T , a membrane segment of linear size L forms spatially anisotropic humps of transverse extension $L_{\perp} \sim (kT/\kappa)L$ (ref. 41). On the other hand, if one applies a lateral tension, Σ , the membrane roughness is strongly reduced and $L_{\perp} \sim (kT/\Sigma)^{1/2}[\ln(L/a)]^{1/2}$, where a is a microscopic length scale.

The shape fluctuations of tensionless fluid membranes represent a boundary case, because the overall gradient of the humps, $L_{\perp}/L$, does not decay for large L . In fact, it follows from the curvature model that the scaling behaviour $L_{\perp} \sim (kT/\kappa)L$ applies only for length scales L that are small compared with the so-called persistence length $\xi_p \equiv a \exp[c\kappa/kT]$, where c is a dimensionless coefficient⁴². Furthermore, as L is increased towards ξ_p , the shape fluctuations reduce the bending rigidity and lead to an effective rigidity $\kappa_{\text{eff}} \approx \kappa - c'kT \ln(L/a)$ in the limit of small kT/κ (refs 43, 44).

For phospholipid bilayers at room temperature, the persistence length ξ_p is usually much larger than the size of the membranes, and the reduction of the bending rigidity by shape fluctuations is negligible. An exceptionally small value for ξ_p is obtained for lecithin membranes containing a small amount of bipolar ('bola') lipid, leading to $\kappa_{\text{eff}} \approx kT$ (ref. 8). Such values for κ_{eff} may also apply to surfactant bilayers that form lamellar and bicontinuous phases^{45–47}.

In general, the membrane should have an effective bending rigidity of the order of kT on length scales $L \approx \xi_p$. It will then undergo rather strong shape fluctuations. The character of these fluctuations is, however, not understood. Real membranes cannot self-intersect and this constraint of self-avoidance is difficult to treat theoretically. Recently, fluid-like membranes with self-avoidance have been investigated by computer simulation, but so far no (bare) bending rigidity has been included⁴⁸.

Three different possibilities can be envisaged for the large-scale behaviour of fluid membranes. (1) The membrane could become highly convoluted or crumpled: a membrane segment of linear size L forms spatially isotropic blobs of size $R \sim L^{\nu}$

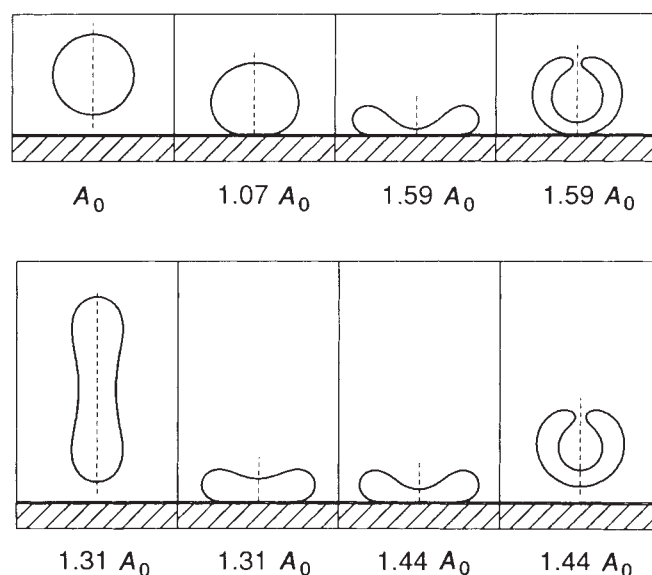


FIG. 3 Shape transformations of a vesicle interacting with a planar surface. These transformations are induced by a change in area at constant volume. The initial state is a sphere with area A_0 . The upper and the lower sequence correspond to two different contact potentials.

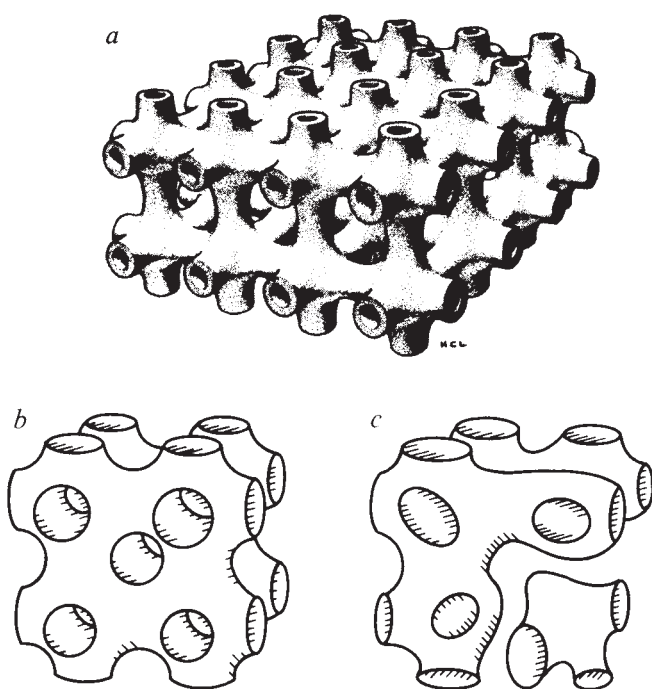


FIG. 4 Bicontinuous structures of bilayers. *a*, Prolamellar body as present in some organelles of plant cells²⁶. *b*, Triply periodic minimal surface. The mean curvature vanishes for each point of this surface. *c*, Sponge phase, obtained by melting of the bilayer crystals shown in *a* and *b*.

with $\nu \geq 0$. A crumpled membrane represents the two-dimensional analogue of a one-dimensional polymer chain^{49,50,37}. In fact, it has been speculated that crumpled fluid membranes exhibit the same scaling properties as branched polymers. This behaviour has been found for one-dimensional ring polymers ('planar vesicles') in two dimensions when deflated by an external pressure⁵¹. If two-dimensional fluid membranes in three dimensions indeed exhibit the same scaling behaviour as branched polymers, the crumpled blobs would scale as $R \sim L^\nu$ with $\nu \approx 1$ (ref. 37). (2) On the other hand, there is no clear evidence so far that self-avoiding fluid membranes governed by bending rigidity are indeed crumpled. Thus, it is possible that self-avoidance prevents crumpling and acts to stabilize the effective bending rigidity at the value $\kappa \approx kT$. (3) If the membrane is allowed to undergo topological changes, it could break up into smaller pieces (Fig. 4c). The energy required to create a spherical vesicle of size ξ_p is $\sim 8\pi kT$, which might be overcome by the translational entropy of the vesicle. Lecithin bilayers containing bipolar lipids might show this behaviour, as they have been seen to expel small vesicles on experimental timescales.⁸

Polymerized membranes. Biological membranes often contain two-dimensional protein networks, such as the spectrin network which is part of the plasma membrane of red blood cells⁵². As the timescales for breaking and reassembling the molecular connections of these networks are usually large compared with the timescales involved in the shape fluctuations⁷⁰, the networks can be regarded as 'fishnets' of fixed connectivity.

The mesh size of protein networks is typically large (~ 100 nm). Polymerized membranes with a much smaller mesh size can be obtained from bilayers of polymerizable lipids^{53,54}. Several types of lipids are now available that have two or more polymerizable units per molecule, polymerization of which can often be accomplished by irradiation of the bilayer with ultraviolet light.

On length scales that are large relative to the mesh size of the network, a polymerized membrane can be regarded as a thin elastic sheet. The shape fluctuations of such a sheet consist of both bending and stretching modes^{55,56}. For a phantom membrane with self-intersections, these shape fluctuations lead to a crumpled state: a membrane segment of linear size L forms blobs of lateral extension $R \sim a[\ln(L/a)]^{1/2}$, where a is a micro-

scopic length scale. For zero bending rigidity, a crumpled state exists for all non-zero temperatures⁵⁰. For finite bending rigidity ($\kappa > 0$), on the other hand, the phantom membrane undergoes a crumpling transition at a critical temperature $T = T_{cr}$ (Fig. 5)⁵⁷⁻⁵⁹. For $T > T_{cr}$, the bending rigidity is irrelevant and the membrane behaves as though $\kappa = 0$; for $T < T_{cr}$, the bending rigidity is relevant and the membrane exhibits a rough but uncrumpled state.

The rough state of a polymerized membrane is again characterized by anisotropic humps: a membrane segment of linear size L makes transverse excursions of size $L_\perp \sim L^\zeta$ (ref. 60). In contrast to fluid membranes with $\zeta = 1$, the roughness exponent ζ now satisfies $\zeta < 1$ and the overall gradient of the hump, $L_\perp/L \sim L^{\zeta-1}$, decays for large L . Computer simulations of

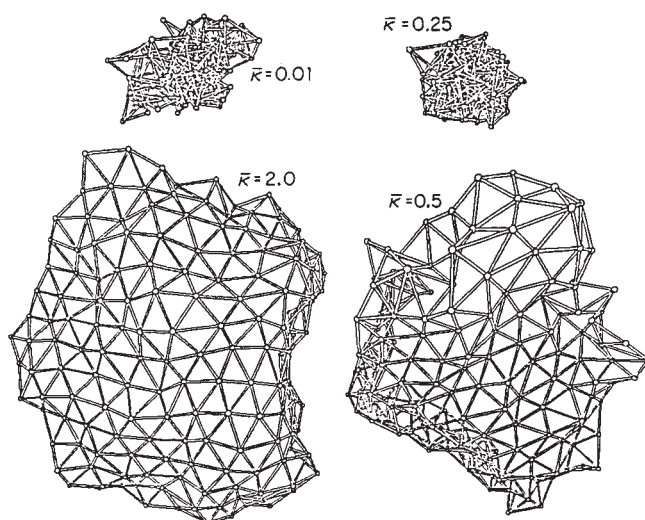


FIG. 5 Crumpling transition of a polymerized membrane with self-intersections, induced by a change in the reduced bending rigidity, $\bar{\kappa} \equiv \kappa/kT$. For $\kappa/kT = 0.01$ and 0.25 , the membrane is crumpled; for $\kappa/kT = 2.0$ and 0.5 , the membrane is rough but not crumpled. The transition takes place at $\kappa/kT \approx 0.33$. (Courtesy of Y. Kantor and D. R. Nelson.)

tethered networks gave an estimate $\zeta = 0.65 \pm 0.05$ (refs 61, 62). More recently, Monte Carlo simulations of solid-like elastic sheets revealed that $\zeta = 1/2$, which is consistent with the existence of a finite shear modulus for the membrane on large scales⁶³.

As mentioned, a real membrane cannot intersect itself. This constraint of self-avoidance does not affect the rough state of a membrane but acts to prevent its crumpling. Recent simulations have shown that the tethered networks considered so far are not crumpled at any finite temperature as a result of self-avoidance^{64–66}.

Crystalline and hexatic membranes. At low temperature or high lateral pressure, monolayers and bilayers of surfactant or lipid molecules exhibit solid-like phases with translational short-range ordering of the molecules. One may often define two different reference lattices: one for the hydrocarbon chains and one for the polar head groups. The mismatch between these two lattices is expected to lead to a high density of defects, such as dislocations⁶⁷.

Another mechanism for the proliferation of dislocations is provided by buckling⁵⁶. For a flat membrane, a single dislocation has a stretching energy that diverges logarithmically with the membrane size. This energy can be reduced by buckling. Recent calculations indicate that, for a buckled membrane, the energy of such a dislocation might be finite⁶⁸. Dislocations would then be thermally activated and thus would melt the crystalline phase at any finite temperature. This could lead to a normal fluid phase, or alternatively to a hexatic phase—a kind of anisotropic fluid⁵⁶.

Theoretically, a hexatic membrane with self-intersections is found to exist at sufficiently low temperatures, characterized by a finite bending rigidity, κ_{eff} , on large scales⁶⁹. Such a membrane is less crumpled than a tethered membrane (with self-intersections). Therefore, hexatic membranes with self-avoidance are presumably not crumpled; their shape fluctuations should then be characterized by the roughness exponent $\zeta = 1$ (ref. 60).

Plasma membrane of red blood cells. On the molecular level, biological membranes are rather complex⁵². One such structure that has been studied extensively is the plasma membrane of red blood cells⁷⁰. This consists of two coupled membranes: a fluid bilayer which contains a mixture of many lipids and proteins, and a two-dimensional network of fibrous spectrin molecules which represents a polymerized membrane (on sufficiently short timescales) with a mesh size of $\sim 0.1\text{--}0.2\text{ }\mu\text{m}$. This network has a small but finite shear modulus, $\mu \approx 5 \times 10^{-6}\text{ J m}^{-2}$ (refs 71, 72).

The flickering of red blood cells has been studied experimentally for a long time^{41,73}. From the theoretical point of view, these fluctuations should exhibit a characteristic crossover from fluid-like behaviour with $L_{\perp} \sim L$ on small scales to solid-like behaviour with $L_{\perp} \sim L^{1/2}$ on large scales⁶³. This crossover should

occur on a scale that is comparable to the mesh size of the spectrin network.

Adhesion and renormalized interactions

Adhesion of membranes plays an essential part in many biological and biophysical phenomena⁵². The formation of tissue, for example, is based on the mutual adhesion of cell membranes or on the adhesion of these membranes to a network of macromolecules. On a somewhat smaller scale, many transport processes involve the binding and unbinding of vesicles to and from the membrane surfaces of cells and organelles. This latter process can be used for the delivery of drugs to specific cell types. The construction of biosensors is often based on the binding of membranes to solid surfaces.

The shape of large adhering vesicles can be observed through a light microscope (Figs 2 and 3). Within the contact region between the vesicle and the substrate, the two surfaces are, on average, parallel and separated by a small gap containing water. They then experience a variety of direct interactions arising from the intermolecular forces.

Direct interaction of membranes. Consider two lipid bilayers in water that carry no electric charge, and that have no macromolecules attached to them. Their direct interaction consists of two parts: (1) a strong short-ranged repulsion usually called the hydration interaction⁷⁴—the molecular mechanism underlying this interaction is still controversial⁷⁵ but its repulsive and short-ranged character is well established; and (2) a longer-ranged van der Waals attraction resulting from the polarizabilities of the water and lipid molecules⁷⁶.

The hydration and van der Waals interactions are 'non-specific'—they are always present for amphiphilic membranes in aqueous solution. In general, the direct interaction includes additional non-specific interactions arising, for example, from electric charges, and 'specific' interactions mediated by biologically relevant macromolecules.

Experiments on membrane interactions. The direct interaction between two membranes can be determined experimentally when these membranes are immobilized on molecularly smooth mica surfaces⁷⁷. The force between two such surfaces can be measured accurately as a function of the surface separation.

The interaction between membranes can also be studied in swelling or dilution experiments of lamellar phases consisting of a stack of membranes. For phospholipid bilayers, the intermembrane spacing has been measured as a function of the external pressure, which can be varied over several orders of magnitude⁷⁸. Typical experimental data resemble the results of Monte Carlo simulations, which are shown in Fig. 6. Swelling experiments have also been performed for stacks of surfactant bilayers^{46,47}. The interaction measured in these swelling experiments, however, represents a renormalized interaction which includes the effect of thermally excited shape fluctuations.

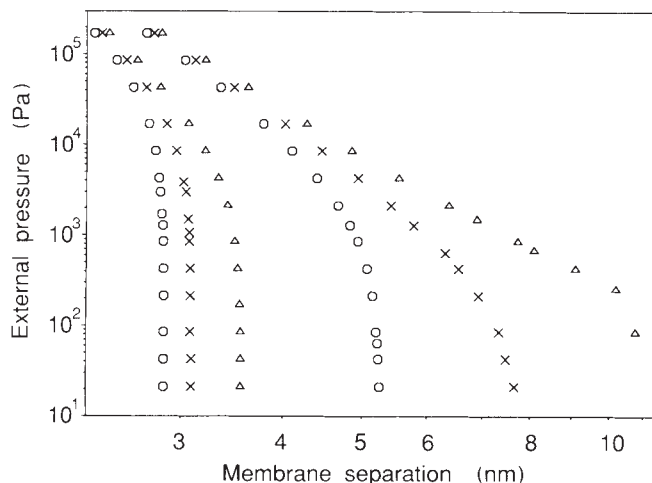


FIG. 6 Monte Carlo simulations of the separation of two interacting fluid membranes as a function of external pressure. The three curves on the right (Hamaker constants $W = 5, 15$ and $25 \times 10^{-21}\text{ J}$) are for room temperature; the three curves on the left represent the behaviour in the absence of shape fluctuations. The difference between these two sets of curves shows the strong renormalization of the direct interaction by shape fluctuations. Δ , $W = 5 \times 10^{-21}\text{ J}$; $\times$, $W = 15 \times 10^{-21}\text{ J}$; $\circ$, $W = 25 \times 10^{-21}\text{ J}$.

Unbinding and adhesion transitions. Thermal shape fluctuations renormalize the direct interaction, increasing its repulsive part^{79,80,37} (see Box B). The renormalized interaction may be attractive or repulsive at large membrane separation, corresponding to a bound or an unbound state of the membranes. These two different states are separated by a phase boundary at which the membranes undergo an unbinding or adhesion transition⁸⁰.

Unbinding transitions were first found theoretically by functional renormalization-group (RG) methods. The RG results imply that two lipid bilayers interacting with realistic intermembrane forces undergo such a transition at a critical unbinding temperature T_u (ref. 80). In addition, the RG calculation predicts that these unbinding transitions are continuous and characterized by universal critical exponents^{60,80-83}. For example, the mean separation, $\langle l \rangle$, of the membrane from the other surface grows continuously as $\langle l \rangle \sim 1/(T_u - T)^\psi$ as the unbinding temperature is approached from below. The critical exponent ψ is universal in the sense that it is independent of the parameters of the direct interaction, and has the exact value $\psi = 1$ for fluid membranes, and the value $\psi \approx 2/3$ (< 1) for polymerized membranes⁶⁰. Similar transitions also occur when a crumpled membrane is adsorbed onto another surface⁸⁴.

Consider two (or several) membranes that are subject to an external pressure and thus have a finite separation $\langle l \rangle$. The presence of an unbinding transition can then be deduced from the behaviour of $\langle l \rangle$ in the limit of vanishing pressure^{80,85}: for $T < T_u$ and $T > T_u$, the membranes attain a finite and an infinite separation, respectively, as the pressure is decreased to zero. This unbinding transition is in many ways analogous to the wetting transition studied extensively in gas-liquid-surface equilibria³⁷.

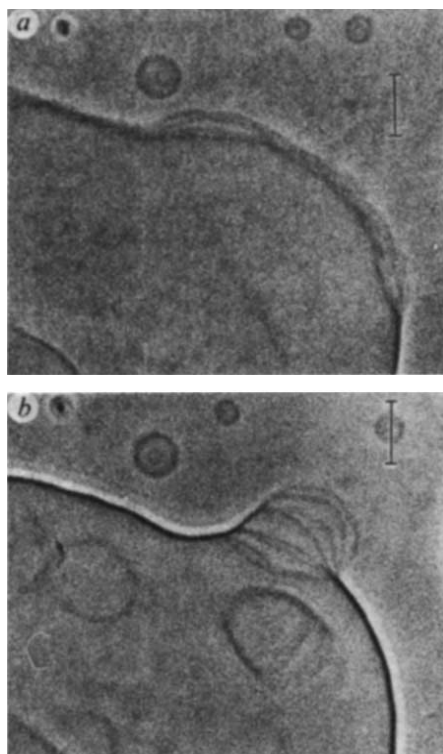


FIG. 7 Unbinding or adhesion transition as observed experimentally for a bunch of lipid bilayers in aqueous solution. *a*, For $T \approx 22.4^\circ\text{C}$, the membranes undulate very strongly, appearing as thick fuzzy lines. *b*, For $T \approx 22.1^\circ\text{C}$, the membranes form a bound state, corresponding to the sharp dark line. The water between the membranes has been squeezed into the large water pocket. The bars represent $10\ \mu\text{m}$. (Courtesy of W. Helfrich.)

For fluid membranes, the existence of unbinding transitions has been confirmed by recent Monte Carlo simulations^{17,86}. The results shown in Fig. 6 were obtained for two membranes with bending rigidities $\kappa_1 = \kappa_2 = 0.2 \times 10^{-19}\text{ J}$ that experience realistic hydration and van der Waals interactions and are subject to an external pressure. The different values of the Hamaker constant W in Fig. 6 determine the strength of the van der Waals interaction. Extrapolation of the room-temperature data gives the critical value $W_u \approx 3 \times 10^{-21}\text{ J}$: for $W < W_u$, the shape fluctuations overcome the van der Waals attraction and the membranes unbind in the limit of zero pressure. Similar results have been obtained for the unbinding transition of solid-like elastic sheets⁶³ and tethered networks (F. F. Abraham and M. Kardar, preprint, MIT).

Unbinding and adhesion transitions of fluid membranes have also been found in recent experiments with sugar-lipid mem-

Box B. FLUCTUATION INDUCED REPULSION AND RENORMALIZED INTERACTIONS

Consider a fluctuating membrane that adheres to another surface. The entropy of such a bound membrane is reduced by the presence of the second surface. All shape fluctuations of the free membrane that exceed a certain wavelength L_{max} are inaccessible to the bound membrane. The difference, $\Delta S = S_b - S_f$, between the entropies of the bound and the free state can be estimated by the difference in the number of accessible modes⁹⁰. For a two-dimensional object, the latter difference scales as $-(R/L_{\text{max}})^2$ for large size, R , of the membrane. The excess free energy per unit area arising from the confinement of the shape fluctuations is then given by

$$V_{\text{fl}} \equiv -kT \Delta S / R^2 \sim kT / L_{\text{max}}^2 \quad (\text{B.1})$$

which represents a fluctuation-induced repulsion acting between the surfaces.

For a rough membrane, the scale invariance of the shape fluctuations is characterized by $L_{\perp} \sim \mathcal{A} L^\zeta$ with roughness exponent ζ . $\mathcal{A}$ depends on temperature and on the elastic moduli. If the membrane is bound to another surface at separation l , this scale invariance holds only up to $L_{\perp} \sim l$ and $L = L_{\text{max}} \sim (l/\mathcal{A})^{1/\zeta}$ with $\tau \equiv 2/\zeta$. It then follows from (B.1) that the fluctuation-induced repulsion is given by³⁷ $V_{\text{fl}} \sim kT \mathcal{A}^\tau / l^\tau$. For fluid membranes with $\tau = 2$ and $\mathcal{A} = (kT/\kappa)^{1/2}$, such a repulsive interaction was first predicted by Helfrich⁷⁹.

Some information about the renormalization of the direct interaction $V(l)$ can be obtained from a simple superposition of $V(l)$ and $V_{\text{fl}}(l)$. To study the properties of unbinding or adhesion transitions, however, one has to use a more systematic theoretical approach which is provided by renormalization-group (RG) methods.

Roughly speaking, the direct interaction $V(l)$ represents the interaction between two surface segments of linear size a , the smallest wavelength available to the shape fluctuations. Within the RG approach, one then calculates the effective interaction, $V(l|t)$, between two segments of linear size $a_t \equiv a \exp(t)$ with $t > 0$. This interaction contains all fluctuations of wavelength L with $a < L < a_t$. As t is increased, one successively includes more shape fluctuations and thus obtains the effective interaction on increasingly larger scales.

The first RG method to be used to calculate the renormalized interaction⁸⁰ was an extension of Wilson's approximate recursion relation⁸¹ and was applied previously to wetting phenomena^{37,92}. This RG scheme leads to the differential flow equation

$$\partial V / \partial t = 2V + \zeta l \partial V / \partial l + \frac{1}{2} v \ln [1 + (a_{\perp}^2 / v) (\partial^2 V / \partial l^2)] \quad (\text{B.2})$$

for the renormalized interaction $V(l|t)$, where the scale factors v and $a_{\perp}$ depend on the temperature and on the microscopic scale a . This RG transformation has a line of fixed points, $V^*(l)$, characterized by $\partial V^* / \partial t = 0$ (refs 60, 81–83). The RG flow in the vicinity of these fixed points then determines the critical behaviour at the unbinding transitions—providing, for example, the critical exponent ψ for the mean separation $\langle l \rangle$.

The above theory does not include any hydrodynamical effects. The fluctuating membrane is, however, coupled to overdamped surface waves in the aqueous medium^{41,93,94}. The amplitude of these waves decays as $\exp(-qz)$ with distance z from the membrane. This is a long-ranged effect which should lead to $L_{\text{max}} \sim l$ and thus, according to (B.1), to an effective repulsion $V_{\text{fl}} \sim T/l^2$ between the surfaces.

branes⁸⁷, in which bunches of membranes were observed by light microscopy (Fig. 7). Above a characteristic unbinding temperature, T_u , these membranes are unbound and exhibit strong undulations, appearing as thick fuzzy lines in Fig. 7a. As the sample is cooled below T_u , the membranes suddenly form a bound state, visible in Fig. 7b as the sharp dark line. Adhesion of the membranes drives the water between them into compact pockets. A similar transition may have been observed in experiments on phospholipid bilayers, in which the swelling behaviour was studied for different levels of salinity⁸⁸.

For $T > T_u$, adhesion of membranes can be induced by a lateral tension, Σ , which acts to suppress the shape fluctuations. Such a tension-induced adhesion has been observed for fluid lecithin membranes⁸⁹. In the limit of small Σ , the membrane separation grows as $\langle l \rangle \sim 1/\Sigma^{1/2}$ which represents a superuniversal scaling law, as it holds for fluid, polymerized and hexatic membranes¹⁷.

Summary and outlook

The conformational behaviour of membranes continues to provide many challenging problems for future research. In the context of vesicle shapes, one poorly understood phenomenon is 'blebbing'—the formation of a long necklace of small vesicles which is expelled from a larger one (refs 4, 95; L. Miao *et al.*, preprint, Simon Fraser Univ.). As far as shape fluctuations are concerned, there is still no clear understanding of the large-scale behaviour of fluid membranes, which may remain uncrumpled as a result of self-avoidance. Likewise, it remains to be seen

whether a realistic model membrane can be found that undergoes a crumpling transition. Experimentally, it is important to probe the whole range of length scales involved in these fluctuations. Some very promising tools are X-ray and neutron scattering. These methods should also be useful for systematic studies of unbinding and adhesion transitions.

In a more general context, another set of puzzles is provided by the very dynamic features of the membrane surfaces of biological cells and organelles. These membranes continuously form and expel small vesicles towards the exterior and the interior of the various compartments⁵². These processes, called respectively budding and endocytosis, look rather similar to the shape transformations of lipid vesicles shown in Fig. 1. Even though budding and endocytosis of biomembranes are typically induced by local changes in the membrane structure, these processes will still fit into the conceptual framework described here if they do not depend crucially on the release of metabolic energy. On the other hand, some form of metabolic energy is certainly involved in the heavy traffic of vesicles which shuttle between different compartments of the cell.

Finally, membranes must have played a crucial role in the origin of life. One may postulate that cellular life began with a vesicle containing just the right mixture of polymers. But where did the amphiphilic molecules that formed the first vesicle come from? □

Reinhard Lipowsky is at the Institut für Festkörperforschung, Forschungszentrum Jülich, 5170 Jülich, Germany.

- Bangham, A. D. & Horne, R. W. *J. molec. Biol.* **8**, 660–668 (1964).
- Lasic, D. D. *Biochem. J.* **256**, 1–11 (1988).
- Sackmann, E., Duwe, H. P. & Engelhardt, H. *Faraday Discuss. Chem. Soc.* **81**, 281–290 (1986).
- Berndl, K., Kaes, J., Lipowsky, R., Sackmann, E. & Seifert, U. *Europhys. Lett.* **13**, 659–664 (1990).
- Canham, P. B. *J. theor. Biol.* **26**, 61–81 (1970).
- Helfrich, W. *Z. Naturforsch.* **28c**, 693–703 (1973).
- Evans, E. *Biophys. J.* **14**, 923–931 (1974).
- Duwe, H. P., Kaes, J. & Sackmann, E. *J. Phys. Fr.* **51**, 945–962 (1990).
- Mutz, M. & Helfrich, W. *J. Phys. Fr.* **51**, 991–1002 (1990).
- Harbich, W. & Helfrich, W. *Z. Naturforsch.* **34a**, 1063–1065 (1979).
- Fromherz, P., Röcker, C. & Ruppel, D. *Faraday Discuss. Chem. Soc.* **81**, 39–48 (1986).
- Deuling, H. J. & Helfrich, W. *J. Physique* **37**, 1335–1345 (1976).
- Svetina, S. & Zeks, B. *Eur. Biophys. J.* **17**, 101–111 (1989).
- Seifert, U., Berndt, K. & Lipowsky, R. *Phys. Rev. A* (submitted).
- Evans, E. & Needham, D. *J. phys. Chem.* **91**, 4219–4228 (1987).
- Seifert, U. & Lipowsky, R. *Phys. Rev. A* **42**, 4768–4771 (1990).
- Lipowsky, R. & Seifert, U. in *Proc. 13th Int. Liquid Crystal Conf.* (Vancouver, 1990).
- Zimmerberg, J. *Biosci. Rep.* **7**, 251–268 (1987).
- Helfrich, W. & Harbich, W. in *Physics of Amphiphilic Layers* (eds Meunier, J., Langevin, D. & Boccarda, N.) 58–63 (Springer, Berlin, 1987).
- Czaja, C., Jekutsch, G., Rothenhäusler, B. & Gaub, H. E. in *Biosensors Int. Workshop 1987* (ed. Schmid, R. D.) 339–340 (VCH Verlag, Weinheim, 1988).
- Cevc, G., Fenzl, W. & Sigl, L. *Science* **249**, 1161–1163 (1990).
- Hauser, H., Gains, N. & Müller, M. *Biochemistry* **22**, 4775–4781 (1983).
- Talmon, Y., Evans, D. F. & Ninham, B. W. *Science* **221**, 1047–1048 (1983).
- Gabriel, N. E. & Roberts, M. F. *Biochemistry* **23**, 4011–4015 (1984).
- Kaler, E. W., Murthy, A. K., Rodriguez, B. E. & Zasadzinski, J. A. N. *Science* **245**, 1371–1374 (1989).
- Gunning, B. E. S. *Protoplasma* **60**, 111–130 (1965).
- Luzzati, V., Tardieu, A., Gulik-Krzywicki, T., Rivas, E. & Reiss-Husson, F. *Nature* **220**, 485–488 (1968).
- Scriven, L. E. *Nature* **263**, 123–125 (1976).
- Harbich, W., Servuss, R. M. & Helfrich, W. *Z. Naturforsch.* **33a**, 1013–1017 (1978).
- Longley, W. & McIntosh, J. *Nature* **303**, 612–614 (1983).
- Dubois-Violette, E. & Pansu, B. (eds) *Geometry and Interfaces Proc. Aussois 1990*, *J. Phys. Colloqu. France* **7** (1990).
- Anderson, D. M., Davis, H. T., Nitsche, J. C. C. & Scriven, L. E. in *Physics of Amphiphilic Layers* (eds Meunier, J., Langevin, D. & Boccarda, N.) 130 (Springer, Berlin 1987).
- Karcher, H. *Manusc. math.* **64**, 291–357 (1989).
- Cates, M. E., Roux, D., Andelman, D., Milner, S. & Safran, S. *Europhys. Lett.* **5**, 733–739 (1988); erratum **7**, 94 (1988).
- Huse, D. & Leibler, S. *J. Phys. Fr.* **49**, 605–621 (1988).
- Porte, G., Appell, J., Bassereau, P. & Marignan, J. *J. Phys. Fr.* **50**, 1335–1347 (1989).
- Lipowsky, R. *Physica Scripta* **T29**, 259–264 (1989); and in *Fundamental Problems in Statistical Mechanics VII* (ed. van Beijeren, H.) 139–170 (North-Holland, Amsterdam, 1990).
- Singer, S. J. & Nicolson, G. L. *Science* **175**, 720–731 (1972).
- Kornberg, R. D. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2564–2568 (1971).
- Clegg, R. M. & Vaz, W. L. C. in *Progress in Protein-Lipid Interactions* (eds by Watts, A. & De Pont, J. J. H. M.) 173–229 (Elsevier, Amsterdam, 1985).
- Brochard, F. & Lennon, J. F. *J. Phys. (Paris)* **36**, 1035–1047 (1975).
- de Gennes, P.-G. & Taupin, C. *J. phys. Chem.* **88**, 2294–2304 (1982).
- Helfrich, W. *J. Physique (Paris)* **46**, 1263–1268 (1985).
- Peliti, L. & Leibler, S. *Phys. Rev. Lett.* **54**, 1690–1693 (1985).
- Di Miglio, J. M., Dvořák, M., Leger, L. & Taupin, C. *Phys. Rev. Lett.* **54**, 1686–1689 (1985).
- Larche, F., Appell, J., Porte, G., Bassereau, P. & Marignan, J. *Phys. Rev. Lett.* **56**, 1700–1703 (1986).
- Safinya, C. R. *et al.* *Phys. Rev. Lett.* **57**, 2718–2721 (1986).
- Baumgärtner, A. & Ho, J.-S. *Phys. Rev. A* **41**, 5747–5750 (1990).
- de Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell University Press, New York, 1979).
- Kantor, Y., Kardar, M. & Nelson, D. R. *Phys. Rev. A* **35**, 3056–3071 (1987).
- Leibler, S., Singh, R. P. & Fisher, M. E. *Phys. Rev. Lett.* **59**, 1989–1992 (1987).
- Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland, New York, 1983).
- Bader, H., Dorn, K., Hupfer, B. & Ringsdorf, H. *Adv. Polymer Sci.* **64**, 1–62 (1985).
- Sackmann, E. *et al.* *Ber. Bunsenges. Phys. Chem.* **89**, 1198–1208 (1985).
- Landau, L. D. & Lifschitz, E. M. *Theory of Elasticity* (Pergamon, New York, 1970).
- Nelson, D. R. & Peliti, L. *J. Physique* **48**, 1085–1092 (1987).
- Kantor, Y. & Nelson, D. R. *Phys. Rev. Lett.* **58**, 2774–2777 (1987).
- David, F. & Gutter, E. *Europhys. Lett.* **5**, 709–713 (1988).
- Nelson, D. R., Piran, T. & Weinberg, S. (eds) *Statistical Mechanics of Membranes and Surfaces* (World Scientific, Singapore, 1988).
- Lipowsky, R. *Europhys. Lett.* **7**, 255–261 (1988).
- Leibler, S. & Maggs, A. *Phys. Rev. Lett.* **63**, 406–409 (1989).
- Abraham, F. F. & Nelson, D. R. *Science* **249**, 393–397 (1990).
- Lipowsky, R. & Girardet, M. *Phys. Rev. Lett.* **65**, 2893–2896 (1990).
- Plischke, M. & Boal, D. *Phys. Rev. A* **38**, 4943–4945 (1988).
- Abraham, F. F., Rudge, W. E. & Plischke, M. *Phys. Rev. Lett.* **62**, 1757–1759 (1989).
- Ho, J.-S. & Baumgärtner, A. *Phys. Rev. Lett.* **63**, 1324 (1989).
- Sackmann, E., Fischer, A. & Frey, W. in *Physics of Amphiphilic Layers* (eds Meunier, J., Langevin, D. & Boccarda, N.) 25–37 (Springer, Berlin, 1987).
- Seung, H. S. & Nelson, D. R. *Phys. Rev. A* **38**, 1005–1018 (1988).
- David, F., Gutter, E. & Peliti, L. *J. Physique* **48**, 2059–2066 (1987).
- Elgsaeter, A., Stokke, B. T., Mikkelsen, A. & Branton, D. *Science* **234**, 1217–1223 (1986).
- Waugh, R. & Evans, E. *Biophys. J.* **26**, 115–131 (1979).
- Engelhardt, H. & Sackmann, E. *Biophys. J.* **54**, 495–508 (1988).
- Zilker, A., Engelhardt, H. & Sackmann, E. *J. Phys. (Paris)* **48**, 2139–2151 (1987).
- Rand, R. P. & Parsegian, V. A. *Biochim. biophys. Acta* **988**, 351–376 (1989).
- Israelachvili, J. N. & Wennerström, H. *Langmuir* **6**, 873–876 (1990).
- Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic, London, 1985).
- Marra, J. & Israelachvili, J. N. *Biochemistry* **24**, 4608–4618 (1985).
- Parsegian, V. A., Fuller, N. & Rand, R. P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2750–2754 (1979).
- Helfrich, W. *Z. Naturforsch.* **33a**, 305–315 (1978).
- Lipowsky, R. & Leibler, S. *Phys. Rev. Lett.* **56**, 2541–2544 (1986); Erratum **59**, 1983 (1987).
- Lipowsky, R. *Phys. Rev. Lett.* **62**, 704–707 (1989).
- Grotehans, S. & Lipowsky, R. *Phys. Rev. A* **41**, 4574–4577 (1990).
- David, F. & Leibler, S. *Phys. Rev. B* **41**, 12926–12929 (1990).
- Lipowsky, R. & Baumgärtner, A. *Phys. Rev. A* **40**, 2078–2081 (1989).
- Leibler, S. & Lipowsky, R. *Phys. Rev. B* **35**, 7004–7009 (1987).
- Lipowsky, R. & Zielinska, B. *Phys. Rev. Lett.* **62**, 1572–1575 (1989).
- Mutz, M. & Helfrich, W. *Phys. Rev. Lett.* **62**, 2881–2884 (1989).
- Lis, L. J., Lis, W. T., Parsegian, V. A. & Rand, R. P. *Biochemistry* **20**, 1771–1777 (1981).
- Servuss, R. M. & Helfrich, W. *J. Phys. France* **50**, 809–827 (1989).
- Frenkel, J. *Kinetic Theory of Liquids* (Clarendon, Oxford, 1946).
- Wilson, K. G. *Phys. Rev. B* **4**, 3184–3205 (1971).
- Lipowsky, R. & Fisher, M. E. *Phys. Rev. B* **36**, 2126–2141 (1987).
- Schneider, M. B., Jenkins, J. T. & Webb, W. W. *J. Phys. (Paris)* **45**, 1457–1472 (1984).
- Milner, S. T. & Safran, S. A. *Phys. Rev. A* **36**, 4371–4379 (1987).
- Evans, E. & Rawicz, W. *Phys. Rev. Lett.* **64**, 2094–2097 (1990).
- do Carmo, M. *Differential Geometry of Curves and Surfaces* (Prentice-Hall, Englewood Cliffs, 1976).

ACKNOWLEDGEMENTS. I thank all my collaborators for enjoyable and fruitful interactions. This work has been partially supported by the Sonderforschungsbereich 266 der Deutschen Forschungsgemeinschaft.

Glass from the Cretaceous/Tertiary boundary in Haiti

Haraldur Sigurdsson*, Steven D'Hondt*, Michael A. Arthur*, Timothy J. Bralower†, James C. Zachos‡, Mickey van Fossen§ & James E. T. Channell§

* Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

† University of North Carolina, Chapel Hill, North Carolina 27599, USA

‡ University of Michigan, Ann Arbor, Michigan 48109, USA

§ University of Florida, Gainesville, Florida 32611, USA

Tektite-like glasses preserved at the Cretaceous/Tertiary boundary at Beloc in Haiti provide clear evidence of an impact event. The glass composition suggests that the impact occurred on a continental shelf region, generating a silica-rich glass with chemical composition that reflects the melting of continental crustal rocks, and a calcium-rich glass produced by the fusion of marl sediments. These findings indicate that catastrophic release to the atmosphere of 10^{15} moles of CO_2 from vaporized marl occurred during the impact.

THE global event that marks the Cretaceous/Tertiary (K/T) boundary has left a striking signature of worldwide mass extinctions of marine biota and a pronounced geochemical anomaly in the sedimentary record, including enrichment of platinum-group elements such as iridium. Hypotheses put forward to account for these phenomena include the impact of a large asteroid on the Earth¹, and a pulse of severe volcanism^{2,3}. Because of poor preservation of K/T boundary sediments, however, neither volcanic nor impact-derived glassy materials, such as tephra or tektites, have hitherto been found to provide unambiguous evidence of the nature of the catastrophic event or the chemical composition of the source region.

The thin band of claystone that marks the K/T boundary in continental and ocean sediments contains microspherules, which have been considered by some to be support of the impact hypothesis⁴⁻⁶, whereas others have interpreted the spherules as authigenic in origin^{3,7,8}. In Colorado and New Mexico, the K/T boundary sediments contain 0.1-mm kaolinite microspherules^{7,8} which coexist with 0.6–0.7-mm shocked quartz minerals in the boundary clay, inconsistent with simultaneous fallout of these two particle types of widely different diameter from the same plume; thus the fallout origin of these small spherules is questionable. It has been suggested that the spherules result from deposition of secondary minerals in vesicles in originally glassy materials⁹ and that the precursor to the K/T boundary claystone was glass¹⁰. Several workers have concluded, on the other hand, that the microspherules and particles are products of authigenic growth, due to diagenetic alteration of glassy material of impact or volcanic origin^{7,8,11,12}. Others have proposed that the spherules are the direct result of impact, formed by solidification of impact melt⁴, but the preservation of the materials has hitherto been too poor to arrive at definitive conclusions. On the basis of geochemical data, the terrain of the impact site has been alleged to be mainly oceanic crust of basaltic composition¹²⁻¹⁴, but a more silicic source has also been suggested⁵.

We document well preserved tektite-like spherules from a glass-rich layer at the K/T boundary in deep-water carbonate sediments in the Beloc region of Haiti, where the boundary deposit forms a layer 10–50 cm thick containing 25% glass spherules from 1 to 6 mm in diameter. The Haiti glasses are

among the earliest preserved glasses on Earth and the oldest known impact glasses. The large size of the spherules and their occurrence in a relatively thick layer may be an indication of the close proximity of this outcrop to the source region. The coarse glass-bearing layer contains shocked quartz grains¹⁵ and an iridium anomaly has also been reported at this level¹⁶.

Stratigraphy

Within the sequence of deep-water carbonates and marls in the Beloc formation on the southern peninsula of Haiti^{14,15}, the K/T boundary interval is clearly marked by the glass-bearing layer. This layer directly overlies uppermost Maastrichtian limestones and chalks of *Abathomphalus mayaroensis* and *Micula murus* foraminiferal and nannofossil zones. It is immediately overlain by 5-cm basal Palaeocene PO (*Guembelitra cretacea*) foraminiferal zone sediments. Within 1.1 m of the top of the glass-bearing unit, typical *Parvularugoglobigerina eugubina* zone taxa first appear (*P. eugubina*, *Eoglobigerina eobuloides*, *Eoglobigerina moskvini*). The nannofossil *Biscutum romeinii* first appears 6.75 m above the glass-bearing horizon, which places the interval above that datum in the *B. romeinii* subzone of the earliest Palaeocene NP1 nannofossil zone^{17,18}. Throughout the lowermost Palaeocene interval at this sequence, both nannofossil and foraminiferal results indicate very high sedimentation rates and a large amount of sedimentary reworking. The presence of an extended *Micula murus* zone (26.5–30 mm) and pre-boundary 29R palaeomagnetic interval (26 m) at the Beloc section suggests that the uppermost Maastrichtian part of the sequence is chronostratigraphically complete on 100-kyr timescales¹⁹⁻²¹.

Petrology of glasses

The glass-bearing rock is pale brown to buff in colour, containing abundant smectite-coated glassy spherules in a silty marl of carbonate (42%) and smectite clay. The spherules make up approximately 25% of the rock, but originally the glass probably constituted about half of the rock, as the smectite is its alteration product. The 1–6-mm-diameter glasses are generally spherical, although ellipsoidal, elongated, tear-drop and bar-bell forms are also present, resembling splash-form tektites (Fig. 1a). They have a very smooth outer shell of light grey to brownish smectite of variable thickness (<1 mm), which also forms smectite vesicle infillings inside the spherules. Beneath the smectite shell, the exposed black glass surface has pits, fine ridges and furrows, probably produced after deposition and during the alteration process from glass to smectite (Fig. 1b). The smooth, spherical smectite shell probably represents the original morphology of the glass spherule. About 2% of the glasses are amber to yellowish in colour and more vesicular than the black glasses. The yellowish glass also occurs as thin (50–150 μm) schlieren or streaks within the black glass (Fig. 1c).

The spherules are mostly massive and entirely crystal-free, consisting of pale brown glass in thin section. The total lack of crystals or microlites in the glass is of great importance with regard to their origin—they thus resemble tektite glasses, which are typically crystal-free, and are unlike volcanic glasses, which

invariably contain either phenocrysts or microlites of high-temperature minerals. This petrographic distinction from volcanic glasses is even more striking when we consider that the andesitic volcanic rocks or magmas that broadly correspond to the Beloc glasses in chemical composition are typically very crystal-rich. This seriously calls into question any fractionation mechanism involving crystal separation as the process responsible for the observed chemical range in the Beloc glasses. The Beloc glasses contain vesicles of diameter 0.05–0.2 mm. In some spherules the vesicles are very rare (<5 vol%) and small, whereas the yellow glasses have up to 30% vesicularity. Some spherules are glass bubbles, consisting of a relatively thin-walled glass sphere filled with smectite. In addition to the smectite, some vesicles are filled by massive crystals of white barytes.

Chemical composition

The glasses range from 44–68% SiO₂ and form well defined compositional trends (Fig. 2). When plotted on mixing diagrams, the glasses define a linear relationship (Fig. 3), consistent with the mixing of two endmembers. The black glassy spherules are the dominant endmember, with about 66–68% SiO₂, enriched in alkalis and slightly enriched in alumina. The other endmember is the rare amber to yellowish glass, Ca-rich (30 wt%) and low in silica (SiO₂ 44%), which also occurs as isolated fragments and as streaks in some black glassy spherules, similar to schlieren commonly observed in tektites (Fig. 1c). The high-Ca endmember is enriched in MgO. As is evident from their high oxide totals (mean 99.85 ± 0.5 wt%), the Beloc glasses are very volatile-poor, in contrast to the hydrous character of volcanic glass

shards found in deep-sea sediments. Unlike the hydrated glass contained in volcanic ash layers in deep-sea sediments, the Beloc glasses have resisted hydration. They resemble tektites in this respect, where the anhydrous nature of the glass renders it a relatively stable structure and impregnable to hydration.

The black Beloc glasses have major- and trace-element composition very similar to the upper continental crust and thus resemble the composition of a wide variety of rock types, including the HNa/K australite tektites²², calc-alkaline andesites, certain deep-sea turbidites²³, argillites, shale and mudrock. The only known entirely glassy rocks of this composition, however, are tektites. The black Beloc glasses have higher alkalis than terrestrial andesites and much higher potassium than HNa/K australites. The MgO content of Beloc glasses is considerably lower than in both australites and terrestrial andesites, and consequently the FeO/MgO ratio in the Beloc glasses is unusually high (1.8–2.4), compared to the others (1.5). Alumina is significantly lower in Beloc glasses than in terrestrial andesites and australites, by ~2 wt%, whereas CaO is somewhat higher. The high-Ca glasses present in the Beloc layer have no chemical or petrographic equivalent among terrestrial igneous rocks, but show some similarities to HNa/K australites and upper Eocene impact microspherules²⁴.

The trace-element distribution in a bulk sample of the black Beloc glasses (Table 1; Fig. 4) is broadly similar to that of the upper continental crust, with a high content of incompatible elements and a strong enrichment in light-rare-earth elements (REE) compared to chondrite (La/Yb = 7.7). HNa/K australites, Ivory Coast tektites and calc-alkaline andesites have

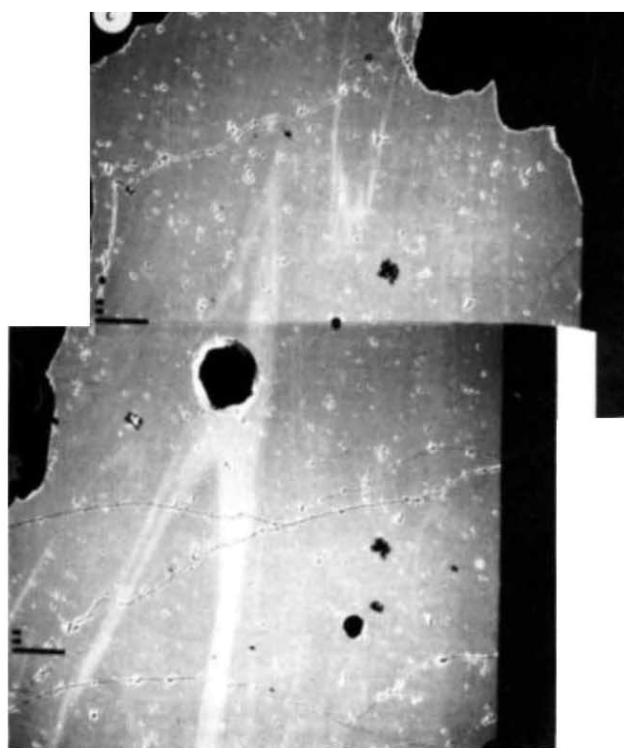
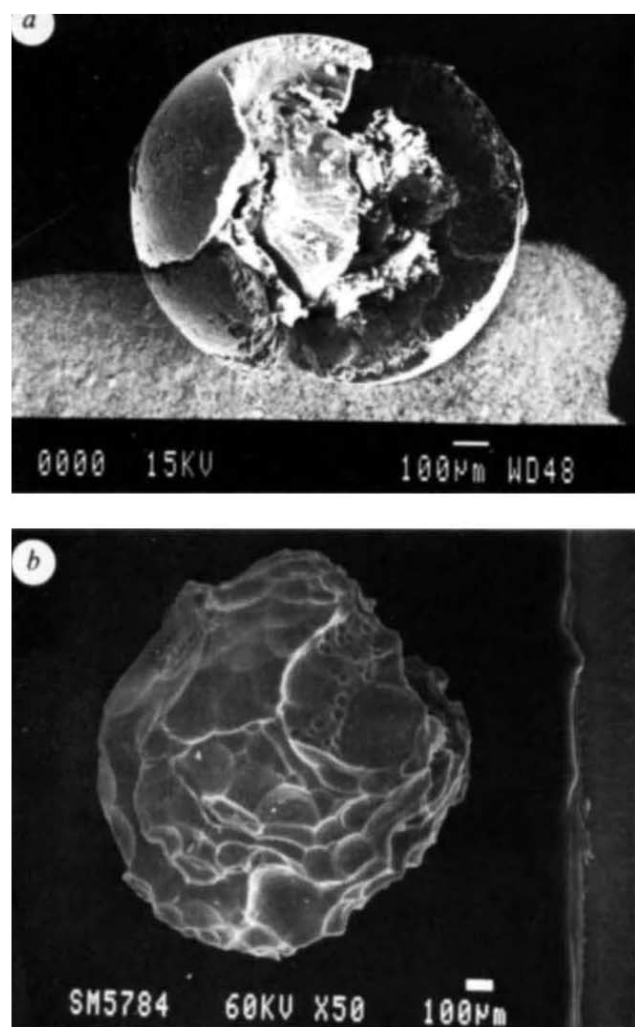
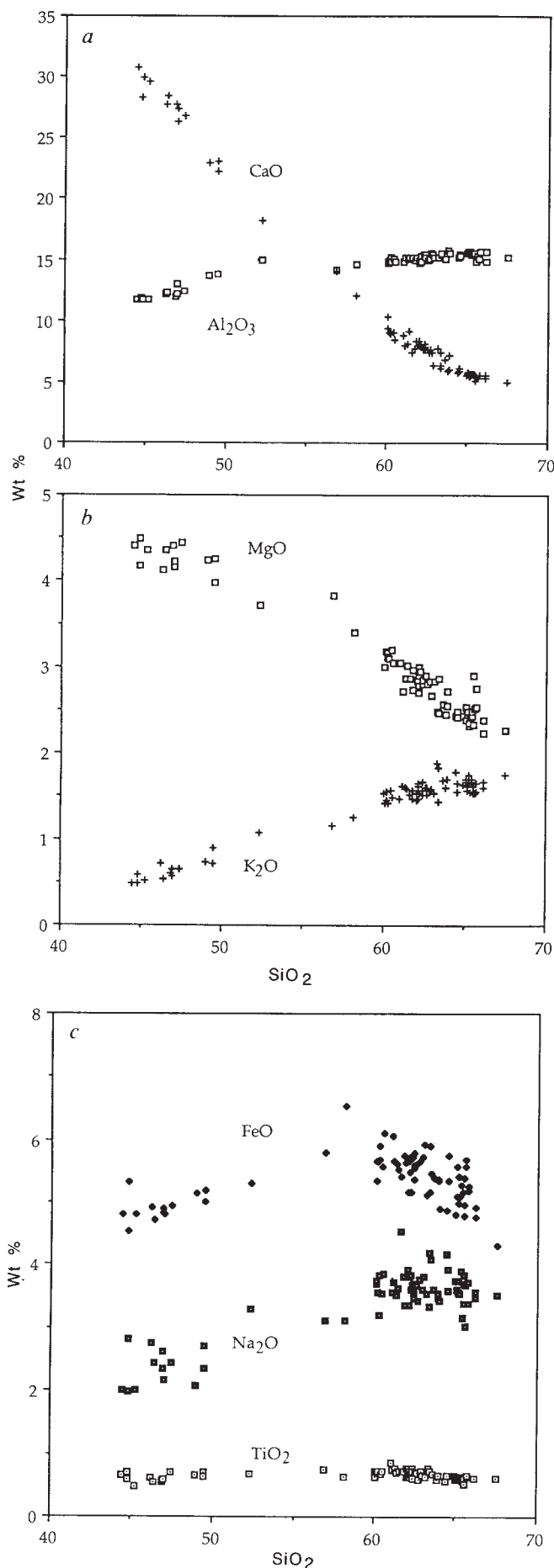


FIG. 1 *a*, Scanning electron micrograph of glass spherule from the K/T boundary at Beloc in Haiti. A smooth smectite shell encloses the glass. *b*, Scanning electron micrograph of glass spherule after removal of the smectite layer. *c*, Backscatter electron micrograph of polished thin section of a Beloc glass spherule. The glass is inhomogeneous, with elongated lighter grey streaks or bands of Ca-rich and silica-poor glass, typically 50–150 μm in width, within the dominant black high-silica glass. Scale bars, 100 μm .



comparable REE patterns^{22,25}. The Beloc glasses have unusually high Th/U ratio (5.8) compared to andesites, however, and are further distinguished from andesites by the very high K/Cs ratio (13,500). The Cs-depletion in the glasses relative to continental crust and volcanic-arc rocks may be an indication of Cs-loss during fractional vaporization from a tektite liquid. The Beloc glasses have rather low Cr and Ni concentrations, comparable to modavite tektites²⁶ and terrestrial andesite²⁷. The Beloc and moldavite tektite glasses thus differ from australites and Ivory Coast tektites, which have Cr and Ni contents that are an order of magnitude higher and which may have been contaminated by the asteroid body that generated these impact melts.

Origin of compositional range

Silicate-liquid compositional trends such as those exhibited by the Beloc glasses could result from fractional crystallization of magma, mixing of two magmas, fractional vaporization of an impact melt at superliquidus temperatures, or the mixing of impact melts derived from a heterogeneous source. A magmatic fractional crystallization origin for the observed trend is excluded because of the total absence of crystals from the glasses and the difference between typical magmatic compositional trends and those of the Beloc glasses. The very high CaO content (up to 30.7 wt%) observed in the Beloc glass low-silica endmember is unknown for volcanic rocks. A magma mixing origin is also discounted on the basis of these observations.

One factor that can lead to compositional diversity among tektite glasses is selective fractional vaporization. The observed order of volatility in silicate melts under conditions of high temperature is $K_2O > SiO_2 > Na_2O > FeO > TiO_2 > MgO > Al_2O_3 > CaO$ ^{22,26,28}. Volatility of alkalis is dependent on oxygen fugacity, and is reduced at atmospheric pressure²⁹. If the compositional range of Beloc glasses were due to fractional vaporization of the melt, then the most silicic and high-alkali glass could represent the closest approximation to the source composition, and the low-alkali and low-silica glass would have suffered greatest fractional vaporization. The observed enrichment and depletion factors of the Beloc glasses are not, however, consistent with fractional vaporization. Whereas CaO shows the increase expected for a refractory oxide, the relative order of enrichment of the other oxides is $K_2O > Na_2O > SiO_2 > Al_2O_3 > TiO_2 > FeO > MgO > CaO$, inconsistent with the oxide sequence predicted for fractional vaporization. Thus Al₂O₃ seems to behave as a slightly volatile oxide, whereas experimental studies clearly demonstrate its refractory behaviour. Furthermore, the degree of fractional vaporization required to account for the trends is excessive. Assuming that Ca is incompatible in the vapour (refractory during volatilization), production of the high-Ca glass would require 85% volatilization of the high-silica glass. Considering the high degree of volatilization implied by the Ca variation, it would indeed be surprising that the alkalis were not depleted to greater extent in the high-Ca glass. Furthermore, as shown in Fig. 3, plots of the ratios of volatile and refractory elements are consistent with mixing and do not show the scatter expected from a fractional vaporization process. We conclude that fractional vaporization had little or no role in producing the observed liquid trend, except in the loss of CO₂.

The glass composition range can, on the other hand, be modelled as a mixing line³⁰ between two endmember compositions. Figure 3 shows that the hypothesis of mixing two rock compositions to produce the glass range is consistent with the observed correlations of log ratios in the Beloc glasses, and that the correlation coefficients are very good. Although the high-Ca and intermediate data points approximate linear compositional

FIG. 2 Oxide variation diagrams for Beloc glasses, including all individual electron microprobe analyses, showing a, CaO and Al₂O₃ plotted against SiO₂, b, MgO and K₂O plotted against SiO₂ and c, FeO* and TiO₂ plotted against SiO₂.

TABLE 1 Composition of Beloc average glass, glass endmembers and calculated 'marl' and 'mudrock' target

	BE4-242 glass*	BE4-242 +CO ₂	Mudrock [†]	Argillite	BE4-42 glass*	Average Beloc glass
SiO ₂	44.44	36.27	62.8	60.5	67.48	63.09 ± 2.13
TiO ₂	0.66	0.54	0.9	0.4	0.61	0.67 ± 0.07
Al ₂ O ₃	11.72	9.56	16.6	18.1	15.14	15.21 ± 0.31
FeO [†]	4.8	3.92	6.8	7.5	4.30	5.44 ± 0.38
MnO	0.1	0.08	0.1	0.1	0.21	0.14 ± 0.05
MgO	4.41	3.6	6.2	4.6	2.26	2.74 ± 0.30
CaO	30.71	25.07	2.4	1.6	4.98	7.26 ± 1.72
Na ₂ O	2.01	1.64	2.8	4.4	3.52	3.63 ± 0.27
K ₂ O	0.48	0.39	0.7	2.5	1.76	1.59 ± 0.12
CO ₂	—	18.59				
Sum	99.32				100.25	99.76

BE4-242 glass, yellow-coloured beloc glass, the rare and most calcic endmember of the glass trend; BE4-242 + CO₂, hypothetical marl composition calculated from calcic glass BE4-242 with CO₂ added to form CaCO₃ with 29 wt% of the CaO present; 'Mudrock', hypothetical 'mudrock' composition calculated by subtraction of CaCO₃ (all CO₂ and stoichiometrically equivalent CaO) from column 2; Argillite, a natural mudrock, from Keweenawan, Canada, renormalized to 100%³¹; BE4-42 glass, black glass, the high-silica endmember of the Beloc glass trend. Average composition of Beloc glasses: mean of 61 electron microprobe analyses, with standard deviation (1σ). Average does not include the rare high-Ca glasses. Trace-element content (p.p.m.) of bulk glass (neutron activation analysis) is: Ni < 85; Cr = 23.7; Co = 13.1; Sc = 19.2; As = 3.3; Br < 0.5; W < 1.5; Sb < 0.18; Rb = 54.0; Cs = 0.97; Ba = 410; Zr = 162; Hf = 3.87; Ta = 0.43; La = 22.3; Ce = 47.7; Nd = 21; Sm = 5.02; Eu = 1.19; Tb = 0.70; Yb = 2.89; Lu = 0.43; Th = 6.35; U = 1.1.

* Electron microprobe analyses.

† Total Fe reported as FeO.

trends, there is significant curvature at the high-SiO₂ end, for example in the CaO trend in Fig. 2a. These deviations from linearity may reflect significant compositional variation within the impact terrain that gave rise to the black glass.

Nature of the impact terrain

Although the mixing model provides a very satisfactory numerical solution, it is not geologically reasonable because the high-CaO endmember is a rock composition not known on Earth. The source terrain that generated this high-CaO melt may have been chemically modified by the melting process, however, due to the loss of a volatile CO₂ component from a geologically realistic rock composition such as marl or interbedded lime-

stones and shales. Thus the most calcic glass could be considered the product of fusion of marl, composed of a mixture of mudrock and CaCO₃, accompanied by loss of CO₂ before quenching to glass. The degassing of CO₂ from the melt may account for the high vesicularity of the Ca-rich yellow glasses. The composition of a 'marl' produced by the mixture of CO₂ and the high-Ca Beloc glass is shown in column 2 of Table 1, after addition of the CO₂ required to balance most of the CaO. The 'mudrock' composition (the rock resulting from removal of the CaCO₃) is shown in column 3, and it may be compared to the mudrock or argillite in column 4. This calculated 'mudrock' is comparable to the composition of common mudrock^{31,32}, with the exception of a slightly higher MgO and a higher Na/K ratio.

Although the melting of a marl source rock can account for the derivation of the high-Ca glass by vaporization of CO₂ and generation of the Beloc glass trend by mixing, the origin of the silicic endmember cannot be attributed to a mudrock source of this composition. The silicic glasses are much lower in MgO than typical mudrock, and have higher Fe/Mg ratio and lower Na/K ratio, as shown in Table 1. The silicic glasses have incompatible element and REE content typical of upper continental crust, such as australites, Ivory Coast tektites, deep-sea turbidites and calc-alkaline andesites, but show uniquely high K/Cs and high Th/U. They could be derived from the complete melting of upper continental crust of broadly granitic composition. The range of glass compositions found at the K/T boundary in Beloc is thus consistent with an impact melt that is dominantly derived from upper continental crustal rocks of granitic or andesitic type, with a minor contribution from the melting of marl. A granitic source is also consistent with the shocked mineral assemblage found in the impact layer at the K/T boundary³³, but the Beloc glass data and shocked mineral assemblage are both inconsistent with the oceanic crustal source of the impact proposed in a number of studies^{12,34}. The glass data are consistent with an impact site on continental crust overlain by marl sediment, such as on a continental shelf or slope. A location of the K/T boundary impact at the Manson impact structure in Iowa has been proposed by several workers^{8,35}. This impact terrain is consistent with the available evidence from the Beloc glasses, as the central uplift of the Manson structure consists of lower and middle Proterozoic granite and gneiss³⁶, overlain by deformed Cretaceous marine shale and limestone^{37,38}. Another

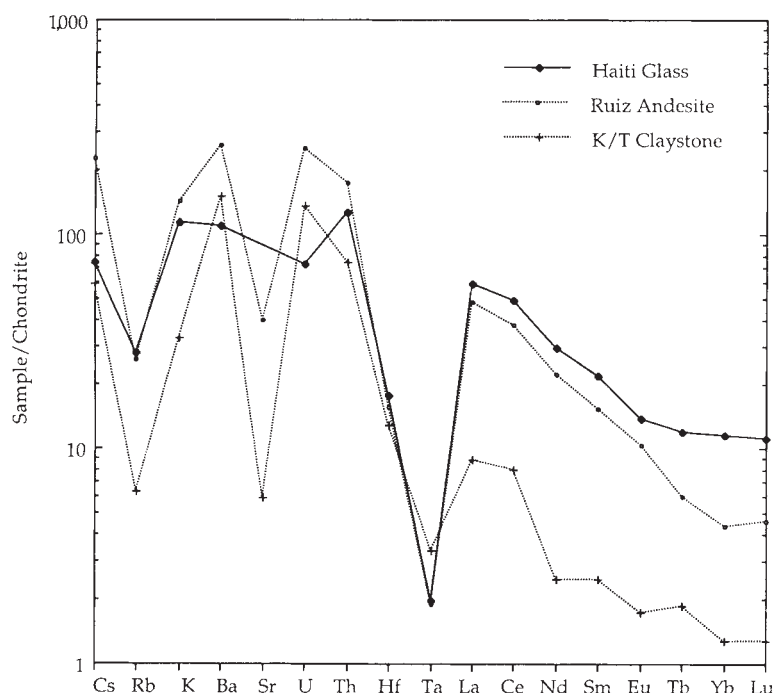


FIG. 3 Logarithmic plots of the oxide ratios of Beloc glasses showing high degree of correlation between a, $\ln(K_2O/CaO)$ and $\ln(SiO_2/MgO)$, and b, between $\ln(MgO/K_2O)$ and $\ln(CaO/Na_2O)$, consistent with an origin of the compositional trend by the mixing of two endmember compositions. The equations of the regression lines are: a, $y = -10.5 + 2.86x$, $r^2 = 0.967$; b, $y = 2.38 \times 10^{-2} - 0.78x$, $r^2 = 0.976$. The glass compositions in middle of the trend occur as streaks in the high-silica glass (Fig. 1c).

possible site of the K/T impact has been proposed on the Yucatan peninsula in Mexico³⁹. We note that the 180-km-diameter Yucatan feature also has a geological succession of the type that could generate the observed Haiti glass compositions, with andesitic rocks, marls and a 80-m-thick bentonitic calcareous breccia. A detailed geochemical comparative study of Beloc glasses and Yucatan and Manson rocks is needed to evaluate these possible impact sites.

The claystone at the K/T boundary in western United States contains common spherules of kaolinite⁸, and the suggestion has been made that the claystone is the product of alteration from glass⁹. Comparison of the chemical composition of Beloc glasses with the K/T boundary claystone are complicated by the mobility of a number of elements during alteration and diagenesis of the claystone. It has been proposed^{8,34} that the Ti/Al ratio in the claystone (0.047 ± 0.005 , 1σ error) reflects that of the progenitor rock and has remained unaffected by diagenesis. The mean Ti/Al ratio in Beloc glasses is 0.050 ± 0.005 (1σ), identical within analytical error. The trace-element distribution in the Beloc glass is compared with that of the K/T boundary claystone in Fig. 4. Although the REE patterns are very similar, the K/T boundary claystone is depleted in all the

rare earths by an order of magnitude compared with the Beloc glasses. A comparable relationship has been observed during the alteration of volcanic glass to clay, where the REEs are strongly depleted overall, but the shape of the chondrite-normalized pattern remains unchanged^{40,41}. The larger cations (Cs, Rb, K) are depleted in the claystone relative to the Beloc glasses, as would be expected during alteration to clay, whereas the content of hafnium, an element that is immobile during alteration of volcanic glass to clay, is identical⁴¹. Finally, both Beloc glasses and the claystone have very low Ni and Cr content, lower than typical shale. The geochemical evidence is thus consistent with a genetic relationship between the K/T boundary claystone and the Beloc tektite-like glass.

If the Beloc glasses are the precursor to the claystone band observed in K/T boundary marine and continental sediments, then the composition of the Haiti glasses can provide clues to the relative mass of impact ejecta and the nature of the volatile fraction released to the atmosphere. Glass constitutes a minimum of 25% of the Beloc layer, but may have constituted up to 58% if the associated smectite is its alteration product. The total glass fraction is a minimum of 0.5 g cm^{-3} , and the yellow glass content is $\sim 2\%$, or 10 mg cm^{-3} . In keeping with the marl hypothesis discussed above, $\sim 23\%$ of the marl (3 mg cm^{-3}) was lost as CO_2 during vaporization; thus the total fraction of marl represented in the impact layer and released to the atmosphere is 13 mg cm^{-3} . If we adopt a 1-cm global mean thickness of the claystone layer as a conservative estimate⁸, the total global glass fallout, based on the mass fraction (500 mg cm^{-3}) in the Beloc sediment, is $2.5 \times 10^{15} \text{ kg}$, corresponding to $\sim 1,000 \text{ km}^3$ or rock. For a bolide mass of $0.6 \times 10^{15} \text{ kg}$ (ref. 1), the impact ejecta mass has been estimated to be 5–200 times the bolide mass⁴², or in the range 10^3 – $4 \times 10^4 \text{ km}^3$. Our minimum estimate of $\sim 1,000 \text{ km}^3$ is at the lower end of this range. The yellow-glass component of the global fallout layer is estimated as $5 \times 10^{13} \text{ kg}$ (20 km^3 rock) on the basis of its abundance in the Beloc section, on a CO_2 -free basis. The CO_2 liberated to the atmosphere from the proposed marl at the time of impact would constitute another 4.6 km^3 . All of the above are minimum estimates, and could be low by a factor of four. Thus we estimate the total 'marl' involved in the impact, as yellow glass in the ejecta layer and as CO_2 in the atmosphere, to be about 25 km^3 . The above calculations imply that about 10^{16} g or $8 \times 10^{14} \text{ moles}$ of CO_2 were emitted to the atmosphere at the time of the impact. This is about three times the estimated annual CO_2 release during Deccan flood-basalt volcanism³ and is comparable to the current annual anthropogenic output. Regardless of the implications for environmental effects from the direct release of CO_2 from impacted material, the glass occurring in the Haiti layer provides direct evidence for a major continental impact at the K/T boundary. □

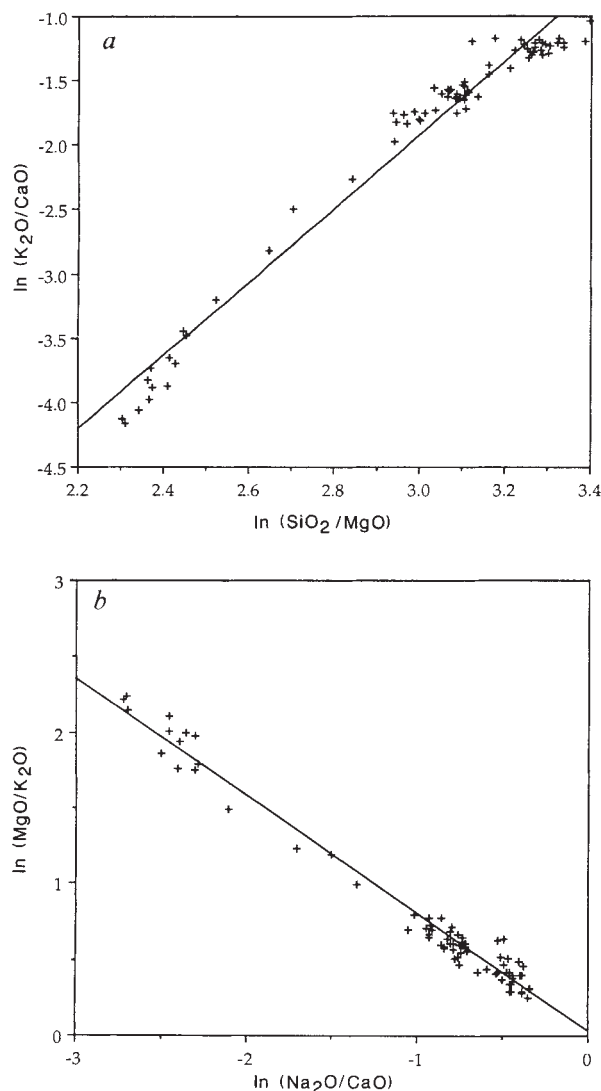


FIG. 4 Chondrite-normalized distribution of trace elements in Beloc glass, compared to K/T boundary claystone from the Raton basin⁸ and Nevado del Ruiz andesite⁴³. The elements are normalized to Leedy chondrite and basalt KD11^{44,45}. The Beloc trace-element data are given in the Table 1 legend.

Received 26 November 1990; accepted 15 January 1991.

1. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
2. McLean, D. M. *Cretac. Res.* **6**, 235–259 (1985).
3. Officer, C. B., Hallam, A., Drake, C. L. & Devine, J. D. *Nature* **326**, 143–149 (1987).
4. Smit, J. & Klaver, G. *Nature* **292**, 47–49 (1981).
5. Smit, J. *Geologie Mijnb.* **69**, 187–204 (1990).
6. Bohor, B. F., Triplehorn, D. M., Nichols, D. J. & Millard, H. T., Jr. *Geology* **15**, 896–899 (1987).
7. Izett, G. A. *Geol. Soc. Am. Bull.* **99**, 78–86 (1987).
8. Izett, G. A. *Spec. Pap. Geol. Soc. Am.* **249**, (1990).
9. Pollastro, R. M., Pillmore, C. L., Tschudy, R. H., Orth, C. J. & Gilmore, J. S. *32nd A. Clay Minerals Conf. Program and Abstracts*, 83 (Clay Minerals Soc., Buffalo, 1983).
10. Pollastro, R. M. & Pillmore, C. L. *J. sedim. Petrol.* **57**, 456–466 (1987).
11. Shaw, H. F. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **60**, 155–177 (1982).
12. DePaolo D. J., Kyte, F. T., Marshall, B. D., O'Neill, J. R. & Smit, J. *Earth planet. Sci. Lett.* **64**, 356–373 (1983).
13. Alvarez, W. *Eos* **71**, 1424 (1990).
14. Hildebrand, A. R. & Boynton, W. V. *Eos* **71**, 1424 (1990).
15. Hildebrand, A. R. & Boynton, W. V. *Science* **248**, 843–847 (1990).
16. Maurasse, F. J. M. R. *Guide to the Field Excursions of Haiti* (Miami Geol. Soc., Miami, 1982).
17. Jiang, M. J. & Gartner, S. *Micropaleontology* **32**, 232–255 (1986).
18. Martini, E. *Proc. 2nd Plankton Conf.* (ed. Farinacci, A.) 736–785 (Edizioni Tecnoscienza, Rome, 1971).
19. Alvarez, W. et al. *Geol. Soc. Am. Bull.* **88**, 367–389 (1977).
20. Berggren, W. A., Kent, D. V. & Flynn, J. J. *Mem. geol. Soc. Lond.* **10**, 141–195 (1985).
21. Monechi, S. & Thierstein, H. R. *Mar. Micropaleont.* **9**, 419–440 (1985).

22. Chapman, D. R. & Scheiber, L. C. *J. geophys. Res.* **74**, 6737–6776 (1969).
23. McLennan, S. M., Taylor, S. R., McCulloch, M. T. & Maynard, J. B. *Geochim. cosmochim. Acta* **54**, 2015–2050 (1990).
24. D'Hondt, S. L., Keller, G. & Stallard, R. F. *Meteoritics* **22**, 61–79 (1987).
25. Schnetzler, C. C., Philpotts, J. A. & Thomas, H. H. *Geochim. cosmochim. Acta* **31**, 1987–1993 (1967).
26. Engelhardt, W. V., Luft, E., Arndt, J., Schock, H. & Weiskirchner, W. *Geochim. cosmochim. Acta* **51**, 1425–1443 (1987).
27. Taylor, S. R. & White, A. J. R. *Bull. Volcan.* **29**, 177–194 (1966).
28. Walter, L. S. & Guirionich, J. E. *Solar Energy* **9**, 163–169 (1967).
29. Walter, L. S. *Geochim. cosmochim. Acta* **31**, 2043–2063 (1967).
30. Delano, J. W. & Lindsley, D. H. *Geochim. cosmochim. Acta* **46**, 2447–2452 (1982).
31. MacPherson, H. G. *Geochim. cosmochim. Acta* **14**, 76 (1958).
32. Blatt, H., Middleton, G. & Murray, R. *Origin of Sedimentary Rocks* 2nd edn (Prentice-Hall, Englewood Cliffs, 1980).
33. Bohor, B. F., Foord, E. E., Modreski, P. J. & Triplehorn, D. M. *Science* **224**, 867–869 (1984).
34. Gilmore, J. S., Knight, J. D., Orth, C. J., Pillmore, C. L. & Tschudy, R. H. *Nature* **307**, 224–228 (1984).
35. French, B. M. *Science* **226**, 353 (1984).
36. Sims, P. K. *U.S. Geol. Surv. Open File Rep.* 85-0604 (1985).
37. Smith, T. A. thesis, Iowa State Univ. (1971).
38. Hartung, J. B. & Anderson, R. R. *Lunar planet. Inst. Tech. Rep.* 88-08 (1988).
39. Hildebrand, A. R. & Penfield, G. T. *Eos* **71**, 1425 (1990).
40. Zielinski, R. A. *Chem. Geol.* **35**, 185–204 (1982).
41. Zielinski, R. A. *Sedimentology* **32**, 567–579 (1985).
42. O'Keefe, J. D. & Ahrens, T. J. *Lunar planet. Sci.* **19**, 883–884 (1988).
43. Melson, W. G. et al. *J. Volcan. geotherm. Res.* **41**, 97–126 (1990).
44. Kay, R. W. & Hubbard, N. *Earth planet. Sci. Lett.* **38**, 95–111 (1978).
45. Kay, S. M. *J. geophys. Res.* **92**, 6173–6189 (1987).

ACKNOWLEDGEMENTS. We thank J. Gardner for determination of the trace-element composition of glasses by neutron activation analysis at the Nuclear Research Center, University of Rhode Island. We thank J. Devine for the determination of major-element composition of glasses by electron microprobe analysis at Brown University.

Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes

Mahshid Company, Jaime Arenas & John Abelson

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

The product of the yeast *PRP22* gene acts late in the splicing of yeast pre-messenger RNA, mediating the release of the spliced mRNA from the spliceosome. The predicted PRP22 protein sequence shares extensive homology with that of PRP2 and PRP16 proteins, which are also involved in nuclear pre-mRNA splicing. The homologous region contains sequence elements characteristic of several demonstrated or putative ATP-dependent RNA helicases. A putative RNA-binding motif originally identified in bacterial ribosomal protein S1 and *Escherichia coli* polynucleotide phosphorylase has also been found in PRP22.

NUCLEAR pre-mRNA splicing takes place on a large particle, the spliceosome^{1–3}, whose function is to fold the pre-mRNA into a substrate for splicing and to catalyse the reaction. The splicing reaction itself consists of two phosphotransfer reactions. In mechanism, it is identical to that of the type II self-splicing introns (for a review see refs 4–7). It is therefore widely believed that pre-mRNA splicing is also an RNA-catalysed reaction. Despite that conviction, ATP hydrolysis is required for most of the discernible steps in spliceosome assembly as well as for each of the two phosphotransfer reactions^{8–14}.

The components of the spliceosome include four small nuclear ribonucleoprotein particles (snRNPs) (U1, U2, U4/U6, and U5) and numerous proteins (for a review see refs 15–21). Many of these proteins have been defined through the isolation of temperature sensitive (*ts*) *prp* mutants (for pre-RNA processing) in *Saccharomyces cerevisiae* (reviewed in refs 19–21). At the non-permissive temperature these mutants block pre-mRNA splicing *in vivo* resulting in various phenotypes. Most accumulate pre-mRNA but at least three of them accumulate the splicing intermediates and two others splice normally but accumulate introns in lariat form. One mutant, *prp22*, the subject of this report, accumulates both intron and pre-mRNA at the nonpermissive temperature.

There are now at least 21 known classes of *prp* mutants (*prp2–prp9*, *prp11*, and *prp16–prp28*). *In vitro* analysis of extracts prepared from a number of *prp* mutants indicates that the PRP gene products are directly involved in the mRNA splicing process^{13,22}. Some are integral snRNP proteins, including PRP8 (ref. 23), which is part of the U5 snRNP and for which a mammalian analogue^{24,25} has been identified, and PRP4, part of the U4/U6 snRNP (refs 26, 27). None of the 8–10 sm core proteins, integral components of each of the four mammalian snRNPs, have yet been identified as *prp* mutants. Other proteins, for example PRP11, are associated with the spliceosome but not specifically with any snRNP (ref. 28). Some PRP proteins, for example, PRP5, PRP6, PRP9, and PRP11, are required at early stages of spliceosome assembly whereas others, for example PRP2, PRP16, PRP17, PRP18 and PRP22, act on the fully assembled spliceosome (see review in ref. 21).

The sequences of most of the *PRP* genes have been determined and in some cases they give a clue to the function of the gene product. We have been especially interested in a set of proteins whose sequences are homologous to a growing group of ATP-dependent RNA helicases epitomized by the eukaryotic protein synthesis initiation factor eIF4A (refs 29–34). Both PRP5 and PRP16 are members of this group^{35,36}. The involvement of these proteins in splicing could explain the role of ATP hydrolysis in spliceosome assembly and in the reaction itself.

Here we demonstrate, through the analysis of splicing in heat-treated extracts, that PRP22 is required for the release of the mRNA product from the spliceosome, a previously unsuspected step. We have cloned and sequenced the *PRP22* gene. The protein sequence is highly homologous to that of PRP16 (ref. 36) and the recently sequenced PRP2 (ref. 37). These three PRP proteins define a novel family of RNA helicase-like genes which mediate three consecutive steps late in the mRNA splicing pathway.

Heat inactivation of PRP22 splicing extract

Heat inactivation of splicing extracts prepared from many *prp* mutant strains has suggested a direct role for the PRP gene products in pre-mRNA splicing²². Extracts were prepared from a *prp22* *ts* strain grown at the permissive temperature (23 °C). Heat lability of splicing activity in this extract was investigated

by pretreatment of the extract at 30–32 °C for 60 min before performing the splicing reaction at 15 °C. Reactions with heat-treated prp22 extract (prp22Δ) showed an increased amount of excised lariat intron (lane i in Fig. 1a) when compared to untreated prp22 extracts or wild-type extracts (lanes c, f, and l in Fig. 1a). This observation mimics the *in vivo* phenotype of the *prp22* mutant at nonpermissive temperatures³⁸. In addition, the total amount of radioactivity detected at 60 min was higher in prp22Δ extracts than in control extracts (compare lane i with lanes c, f, and l in Fig. 1a). This observation is reproducible and is most probably due to increased stability of the pre-mRNA and its products in the prp22Δ extract (see the following section for further explanation). Addition of heat-inactivated extracts from prp5, prp11, or prp18 to the prp22Δ extract restores amounts of lariat intron to wild-type levels (data not shown), suggesting that intron accumulation in prp22Δ extract is due to specific inactivation of an exchangeable factor(s) (see Lustig *et al.*²² for detailed arguments about complementation analysis).

PRP22 gene product acts late in splicing

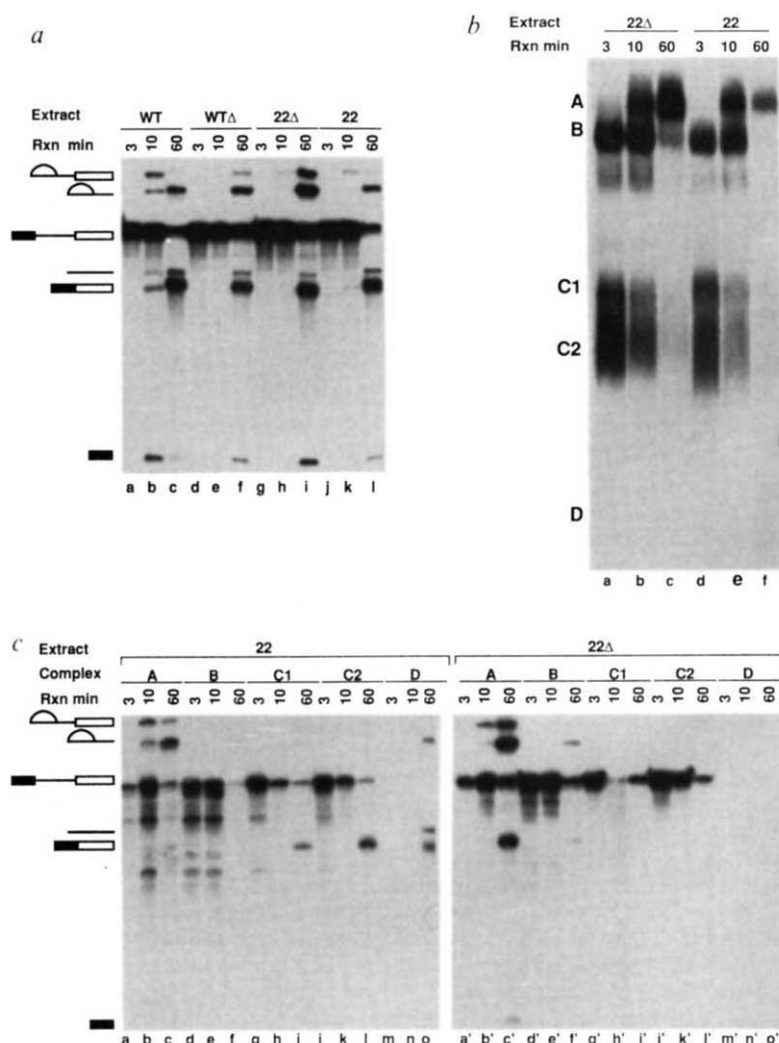
To determine the step in the splicing pathway at which the PRP22 gene product acts, complexes formed in prp22Δ extracts were investigated by electrophoresis as described by Cheng and Abelson⁹. Five complexes were detected: 'A', the active spliceosome corresponding to complexes A1, A2-1, and A2-2 in Cheng and Abelson, and B, C1, C2, and D. The complexes corresponding to C1 and C2 were also detected by Cheng and Abelson but they were not further investigated. Complex D was not detected

by Cheng and Abelson as a result of prolonged electrophoresis on a shorter gel.

All five complexes had similar patterns of formation in heat-treated and untreated wild-type or prp22 extracts (wild-type data not shown). Complex D, however, was not detectable in the prp22Δ extract (See Fig. 1b). In addition, there was a greater amount of complex A in prp22Δ extracts after 60 min compared to controls at 60 min (approximately 2–3-fold increase in the amount of radioactivity, data not shown).

To identify the RNA components in each complex, RNA from complexes A, B, C1, C2, and D were eluted from the gel and analysed by denaturing polyacrylamide gel electrophoresis (Fig. 1c). Complex B formed early during the splicing reaction and contained predominately pre-mRNA in all extracts at all time points (lanes d–f and d'–f' in Fig. 1c). The small amounts of lariat intron–exon 2 and lariat intron detected may be due to contamination from complex A. At early time points, complex A contained only pre-mRNA (lanes a and a' in Fig. 1c), but later (10 and 60 min) contained excised lariat intron, exon 1, lariat intron–exon 2, pre-mRNA, and very small amounts of spliced mRNA in WT, WTΔ (data not shown) and prp22 extracts (lanes b and c in Fig. 1c). In contrast, complex A in the prp22Δ extract contained a much larger amount of spliced mRNA in addition to an increased level of lariat intron at 60 min and further, these two RNA species, mRNA and lariat intron, were present in about equal molar amounts (lane c' in Fig. 1c). Complexes C1 and C2 also formed very early in the splicing reaction and contained pre-mRNA (lanes g, h, j, k, g', h', j' and

FIG. 1 *a*, Heat inactivation of prp22 extracts. Lanes a–c, untreated wild-type (WT) extracts; lanes d–f, heat-treated wild type (WTΔ) extracts; lanes g–i, prp22 heat-treated (22Δ) extracts; lane j–l, untreated prp22 (22) extracts. RNA species are indicated to the left of the gels by the following symbols: solid box, exon 1; open box, exon 2; thin line, intron. Splicing extracts were prepared from the wild-type strain, SS330 (ref. 38), and the haploid *prp22* ts strain grown at permissive temperature (23 °C) according to Lin *et al.*¹¹. Actin precursor RNA was synthesized *in vitro* as described by Cheng and Abelson⁹. Splicing extracts were heat inactivated at 30–32 °C for 60–70 min in the presence of 2 mM MgCl₂, 0.8 mM DTT, and 5 units of RNasin for each 3 μl of splicing extract. Following heat treatment, 10.5 μl of splicing reaction mix (3.6 mM MgCl₂, 1.4 mM DTT, 8.5 mM ATP, 2.8 mM spermidine, 143 mM K₂HPO₄ (pH 7.0), 2.8 mM EDTA, 7% PEG-8000, 20,000–40,000 c.p.m. (~5 fmol) of labelled actin precursor RNA) was added to 19.5 μl heat-treated splicing extract. The splicing reaction was carried out at 15 °C, and samples of 8 μl were taken at each time point. Half (4 μl) of each sample was analysed on denaturing polyacrylamide gels as described by Lin *et al.*¹¹. *b*, Formation of splicing complexes. The remaining halves of the samples taken in *a* were analysed for complex formation on native polyacrylamide gels according to Cheng and Abelson⁹ except that a larger gel (24 × 18 × 0.8 cm) was used. The positions of complexes A, B, C1, C2, and D are indicated. Visualization of complex D in lane f required longer exposure of the film (not shown). *c*, RNA contents of the complexes. The complexes formed in *b* were excised from the gel and RNA species were eluted from the gel slices and analysed on denaturing polyacrylamide gel as in *a*.



k' in Fig. 1c). At early times these are probably nonspecific complexes and can form in the presence of high concentration of heparin which completely blocks spliceosome formation (our unpublished observation). However, at later time points in WT, WTΔ and *prp22* extracts, other complexes containing predominantly the spliced mRNA comigrated at the same position in the C1 and especially the C2 region (lanes i and l in Fig. 1c). In contrast, the *prp22Δ* extract did not contain any spliced mRNA in the C1 and C2 region at later time points (lanes i' and l' in Fig. 1c). Complex D formed late in the splicing reaction and contained a low level of both spliced mRNA and intron (lane o in Fig. 1c). As mentioned earlier, this complex was not detectable in the *prp22Δ* reactions (lane o' in Fig. 1c).

These observations indicate that in the wild-type splicing extracts, the spliced mRNA leaves the spliceosome soon after ligation of the exons, appearing in the fast-migrating complexes in the regions of C1 and C2, and in complex D. This is consistent with earlier observations that spliced mRNA is in a low molecular weight complex^{10,39}. In *prp22Δ* extracts, however, the spliced mRNA apparently remains associated with the spliceosome (complex A) and is not released into the low molecular weight complexes. This suggests that the PRP22 gene product plays a part in the release of spliced mRNA from the spliceosome.

In addition to retention of mRNA in the spliceosome in *prp22Δ* extracts, a larger amount of lariat intron is detected in complex A when compared with controls. Previous *in vitro* analyses of excised lariat intron have indicated that it is not released as quickly as spliced mRNA from the spliceosome and is found associated with snRNPs in a high molecular weight particle^{10,39}. But eventually the lariat intron is thought to dissociate from the snRNPs, after which it is debranched, and finally degraded. The observation of large amounts of lariat intron in the *prp22Δ* splicing extract may indicate that it has been stabilized in the spliceosome as a result of the defect caused by the *prp22* mutation.

On the basis of the accumulation of mRNA and intron in the spliceosome, we suggest that the PRP22 gene product must act early in spliceosome dissociation. A defect in this factor may result in a defective spliceosome incapable of releasing the products of the splicing reaction and the components associated with it. As a consequence, these components cannot be recycled into new rounds of pre-mRNA splicing, which results in pre-mRNA accumulation in the nucleus. Also, because of its retention in the defective spliceosome, mRNA cannot be transported to the cytoplasm. This eventually results in the death of the cell owing to depletion of the cytoplasmic pool of processed mRNA.

In summary, the *in vitro* data presented here indicates that

the PRP22 gene product is required at the final stages of the splicing process after the splicing reaction itself is completed. This gene product is involved either directly or indirectly in release of spliced mRNA from the spliceosome.

Isolation of the PRP22 gene

We cloned the wild-type PRP22 gene by complementing the *ts* phenotype of the *prp22* strain with plasmid DNA isolated from a YCp50-yeast genomic DNA library (see ref. 13). Plasmid DNAs from temperature-resistant transformants were isolated and reintroduced into the *prp22 ts* strain. In all cases these DNAs conferred high temperature resistance (37 °C) on the *prp22 ts* strain. Analysis of PRP22 clones by subcloning various fragments narrowed the complementation activity to an *EcoRV*-*NruI* DNA fragment of ~4 kilobases (kb) (Fig. 2a).

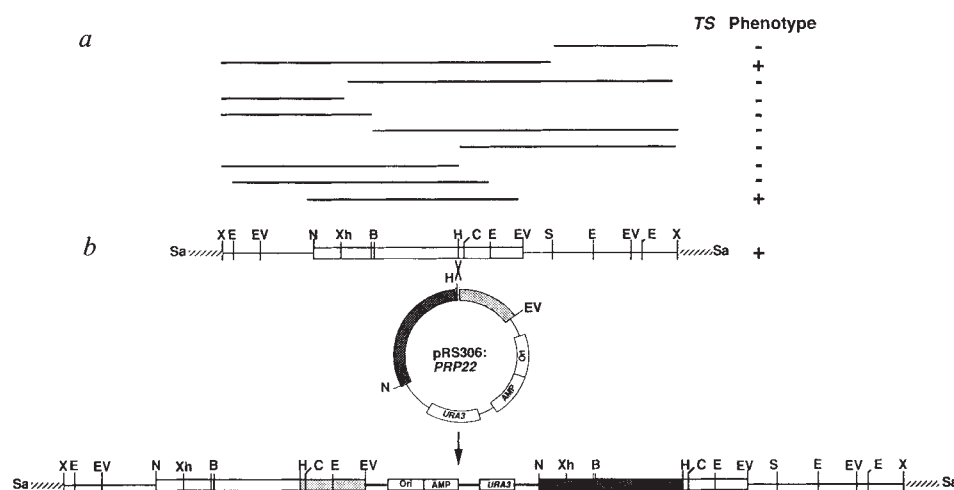
Genetic evidence that the complementing clone harbored the wild-type PRP22 allele rather than an extragenic suppressor was provided by mapping the cloned PRP22 DNA in transformants in which a PRP22-bearing plasmid had integrated into the chromosomal region homologous to the cloned insert. To do this, the *EcoRV*-*NruI* complementing fragment was subcloned into a yeast integrative plasmid, pRS306 (ref. 40). To direct the integration of the subcloned pRS306 plasmid into the chromosomal region, plasmid DNA was cleaved at the unique *HpaI* restriction site in the putative PRP22 gene (Fig. 2b). The linear plasmids were then used to transform the TS⁺ parental yeast strain (SS330) to *URA*⁺ prototrophy. *URA*⁺ transformants were isolated and were crossed to a *prp22 ts ura3* haploid and then tetrads (3–10 tetrads) from the sporulated diploids (5 independent crosses) were analysed. The cosegregation of the *ts* and *ura3* phenotype and of the *URA3* and TS phenotype in the resultant tetrads demonstrated that the cloned DNA was integrated at the PRP22 chromosomal locus. Southern analysis of the *URA*⁺ transformants also confirmed that the subcloned pRS306 plasmid had integrated at the genomic PRP22 locus (data not shown).

PRP22 gene sequence

The DNA sequence of the ~4.0-kb *EcoRV*-*NruI* fragment was determined. The sequence revealed a 3,435-nucleotide open reading frame which potentially can encode a polypeptide of 1,145 amino acid residues with a predicted relative molecular mass ~130,000 (Fig. 3). Homology searches in different data banks using the FASTA and TFASTA algorithms of Pearson and Lipman⁴¹ revealed that the PRP22 protein shares remarkable homology with several different proteins described below.

Sequence similarities with ribosomal protein S1 and polynucleotide phosphorylase. Sequence analysis of the N-terminal region

FIG. 2 a, *In vivo* complementation analysis with the clone and subclones of PRP22. A partial restriction map of a *sau3A* genomic clone and the phenotypes of the subclones are shown. Horizontal bars indicate the restriction fragments analysed for their ability to complement the *ts* phenotype of a *prp22 ts* strain. Ability to complement the *ts* defect is indicated by '+'; inability by '-'. Restriction enzymes: B, *Bam*HI; C, *Cl*AI; E, *Eco*RI; EV, *Eco*RV; H, *Hpa*I; N, *Nru*I; S, *Sac*I; Sa, *Sau*3A; X, *Xba*I; Xh, *Xho*I. Complementation analyses were done as described by Vijayraghavan and Abelson¹³. b, Diagram of the integration event of the cloned DNA at the PRP22 locus. The open box represents the chromosomal locus. Cloned DNA is shown by the dotted box. The boxes in the plasmid are labelled: AMP, ampicillin resistance; Ori, ColE1 origin of replication; *URA3*, a yeast marker. Other genomic or plasmid DNA sequences are represented by thin lines.



of PRP22 revealed the presence of a stretch of 70 amino acids which was very similar to a motif present in the bacterial ribosomal protein S1 (RPS1)^{42,43} and *E. coli* polynucleotide phosphorylase (PNPase)⁴⁴ (Fig. 4). The function of this motif is unknown. Four tandem repeats of the motif are found at the central and the C-terminal regions of RPS1 and it has been suggested to have RNA-binding activity (see reviews in refs 43 and 45).

Using the PROFILESEARCH program of Gribskov *et al.*⁴⁶, sequences related to the S1 motif were also revealed in eukaryotic translation initiation factor-2 α subunit (eIF2 α) as well as *E. coli* and chloroplast initiation factor-1 (IF1) (Fig. 4). In contrast

to the ribosomal S1 protein, PRP22, PNPase, IF1, and eIF2 α contain only one copy of the S1 motif. Its relative position in each protein does not seem to be important (Fig. 4b). The IF1 protein is only 72 amino acids long⁴⁷, suggesting that this motif by itself may constitute a complete functional domain.

Sequence similarities with PRP2, PRP16, and other helicases. Our searches also revealed that PRP22 shared extensive sequence homology with two other PRP proteins, PRP16 (ref. 36) (or PRP23; *prp23* (ref. 38) is allelic to *PRP16*) and PRP2 (ref. 37), in a region of around 700 amino-acid residues located at their central and C-terminal regions (Fig. 5). PRP2 is an extrinsic factor in splicing and is required for the first step of

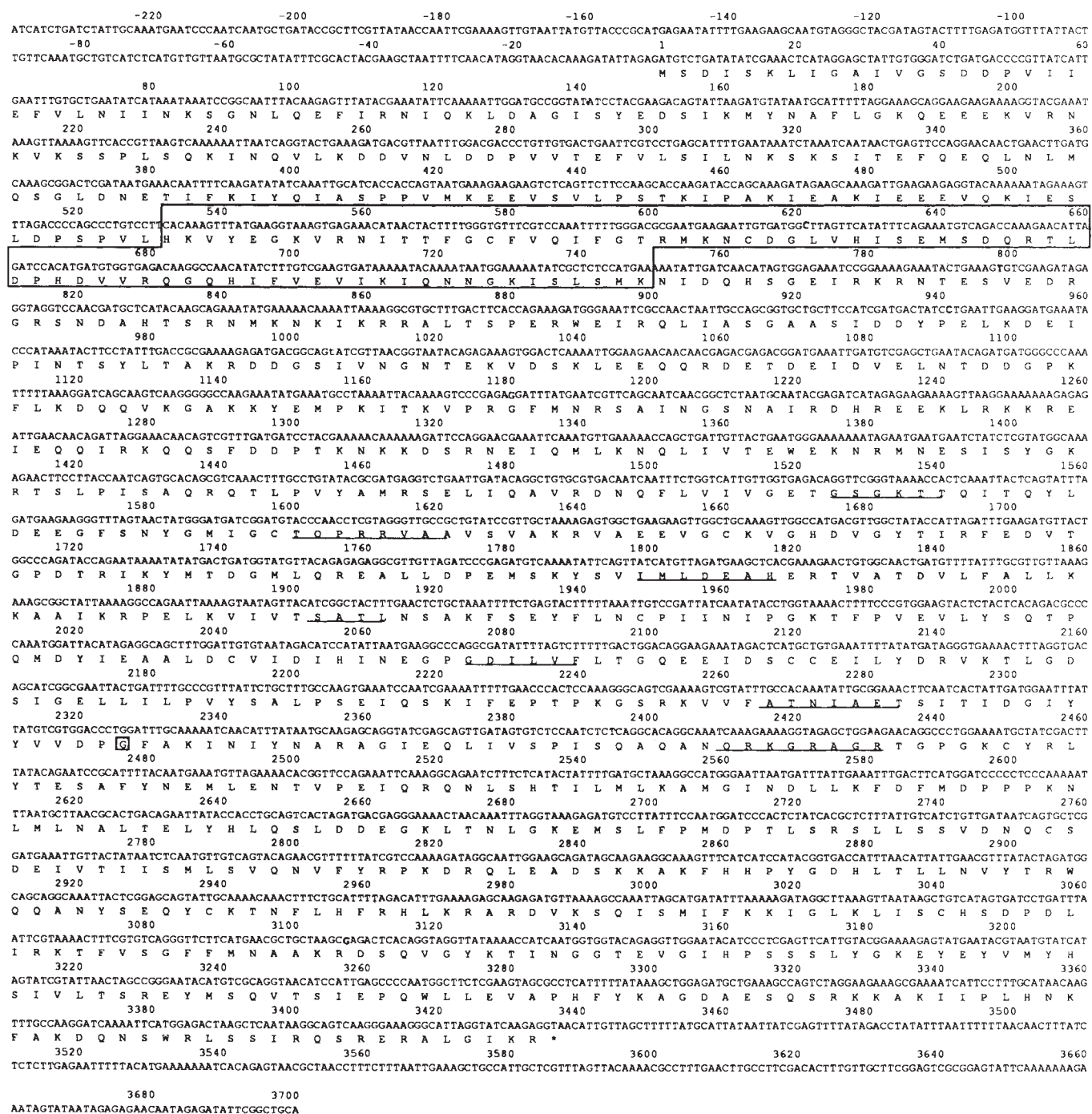


FIG. 3 Nucleic acid and deduced amino acid sequence (single-letter amino acid code) of the *EcoRV-NruI* fragment that complements the *ts* defect of *prp22*. The region with similarity to bacterial ribosomal protein S1 and *E.*

coli polynucleotide phosphorylase is boxed. Regions of similarity with seven conserved elements found in RNA helicases are underlined. The wild-type glycine residue affected by the *prp22* mutation is boxed.

the splicing reaction¹¹. The *PRP16* gene was originally isolated as a suppressor of a branchpoint mutation (an A to C mutation in the last A in the conserved TACTAAC box) which inhibits splicing at the first step⁴⁸.

The central domain (residues 480–840) of PRP22 shows 55 and 52% identity and 72 and 74% similarity with the corresponding domains of PRP2 and PRP16, respectively. Previous analysis of the PRP16 protein revealed that the central region of this protein contains an NTP-binding motif and is similar to some of the sequence motifs characteristic of RNA-dependent ATPases and helicases³⁶. We therefore compared the central region of these three PRP proteins to different families of putative helicases. It shows homology to RNA helicase-like proteins encoded by three different families of positive-strand RNA viruses (flaviviruses, pestiviruses, and potyviruses) and to the members of the 'DEAD' family of RNA helicase-like proteins. The homologies lie in segments that include seven highly conserved motifs described for a number of putative helicase families (shown by roman numerals in Fig. 5)^{29–31,33,49}. The homology within each motif is well conserved in the three PRP proteins but less so when compared to the RNA helicase-like proteins from the DEAD or viral families. Motif I (GXGKT, single letter amino-acid code) corresponds to the A-site of the nucleoside triphosphate (NTP)-binding motif^{29,32,50} and is found in most nucleotide-binding proteins including ATPases, kinases, and DNA and RNA helicases. Motif II corresponds to the B site of the NTP-binding motif which interacts through the invariant 'D' residue with the Mg²⁺ moiety of Mg-ATP (ref. 51).

There seems to be a correlation of sequence content between motif II and motif VI when different helicase families are compared. If the consensus of motif II is DEAD, the corresponding motif VI will be HXXGRXXR. But when motif II is DEXH, the sequence of motif VI is more likely to be QXXGRXXR (ref. 31). Motif VI is thought to be a nucleic acid binding site because it contains positively charged residues.

ATP-dependent RNA helicase activity has been shown *in vitro* for only two members of the DEAD family, eIF4A and P68 (refs 52–54). Because of the extensive sequence similarities

among all members of the DEAD family, there has been speculation that all members of this family have RNA helicase activity. Recently, RNA helicase and RNA-dependent ATPase activities have been demonstrated in the cylindrical inclusion protein of plum pox virus (PPV-CI)^{55,56}. In addition, RNA-dependent ATPase activity has recently been demonstrated for the PRP16 protein¹⁴.

Although PRP22, PRP16 and PRP2 all contain the seven conserved sequence motifs, they cannot be classified as members of DEAD or the viral families of RNA helicase-like proteins. Thus we propose that these PRP proteins define a novel family of RNA-helicase-like proteins which we call the 'DEAH' family on the basis of the sequence of motif II. Further investigation is, however, required to show that these three PRP proteins are indeed RNA helicases. The ts phenotype of *prp22* results from a single amino-acid change within the putative RNA helicase domain. The mutant *prp22* gene contains a G to A change at nucleotide position 2,327 which results in the replacement of a conserved glycine with a glutamic acid residue (Fig. 3). This mutation does not, however, fall within any of the seven highly conserved motifs. It will be interesting to find out how this mutation results in the ts phenotype of PRP22.

As described earlier, PRP22 also shares extensive sequence homology with PRP2 and PRP16 at its C terminus (amino acids 840–1,145). In this region, the PRP22 protein shows 31 and 45% identity and 57 and 67% similarity with PRP2 and PRP16 proteins, respectively. This domain is unique among these three PRP proteins and does not share any homology with other proteins. The extensive sequence homology observed among the three PRP proteins in their C-terminal regions may imply that this homologous domain functions similarly in splicing.

New members of the DEAH family

Because of the high degree of homology between PRP16, PRP22 and PRP2, and the homology with other helicase-like proteins, we decided to search for additional members of the DEAH family in yeast. Degenerate oligonucleotide primers corresponding to box I, box II, and box V were used to amplify yeast

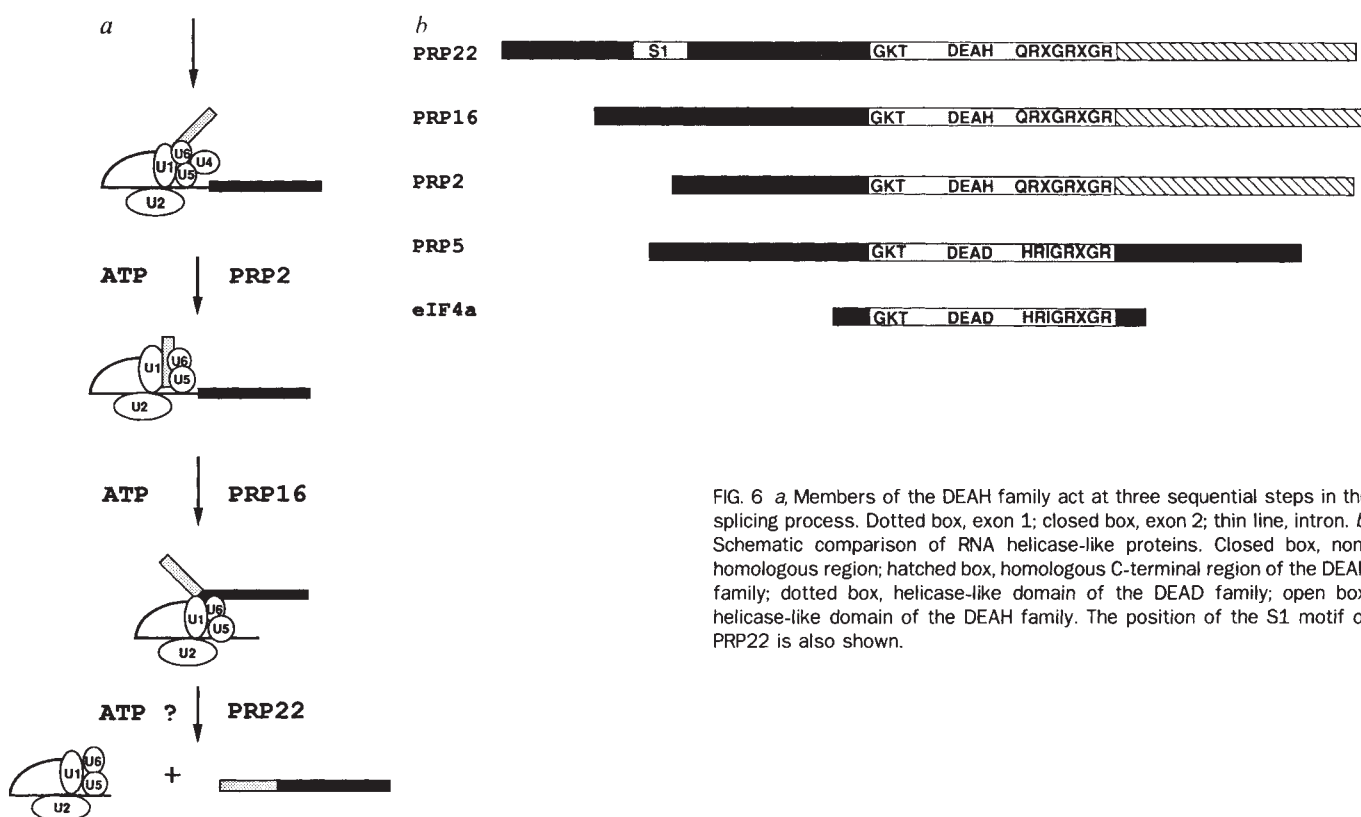


FIG. 6 *a*, Members of the DEAH family act at three sequential steps in the splicing process. Dotted box, exon 1; closed box, exon 2; thin line, intron. *b*, Schematic comparison of RNA helicase-like proteins. Closed box, non-homologous region; hatched box, homologous C-terminal region of the DEAH family; dotted box, helicase-like domain of the DEAD family; open box, helicase-like domain of the DEAH family. The position of the S1 motif of PRP22 is also shown.

genomic DNA by the polymerase chain reaction (PCR). Among the clones analysed, two new members of the DEAH family, JA1 and JA2, were identified in addition to clones corresponding to PRP2, PRP22, and PRP16. The amino-acid sequence alignment of JA1 and JA2 with the other three PRP proteins is shown in Fig. 5. It will be interesting to know whether these two new members are also involved in pre-mRNA splicing as well. The exciting identification of these two new members of the DEAH family suggests that there may be additional members which we have failed to detect here. In addition, the presence of this novel family in other organisms needs to be investigated.

Functional implications

Pre-mRNA splicing requires ATP at several steps during assembly of the spliceosome, and for the first and the second step of the splicing reaction⁸⁻¹⁴. An ATP requirement for the steps following exon ligation, however, has not yet been demonstrated. Identification of NTP-binding motifs in five helicase-like PRP proteins (PRP2, PRP5, PRP16, PRP22, and PRP28) explains the requirement of ATP in the splicing process. PRP5 (ref. 35) and PRP28 (E. J. Strauss and C. Guthrie, submitted for publication) belong to the DEAD family of RNA helicases and are required before the first step of the splicing reaction. The PRP5 protein is required for spliceosome assembly¹¹. In contrast, the members of the DEAH family (PRP2, PRP16, and PRP22) act at three consecutive steps in splicing after the formation of the spliceosome (Fig. 6a). The PRP2 protein is required for the first step of the splicing reaction¹¹, PRP16 acts at the second step¹⁴ and finally, PRP22 is required for the release of message from the spliceosome.

During mRNA splicing there are a number of potential RNA-RNA interactions (intra- or intermolecular) in pre-mRNA and snRNAs. These include U4 pairing with U6, U1 with the 5' splice site, and U2 with the branchpoint sequence (see refs 16 and 17). ATP-dependent RNA helicases may be required to unwind some of these interactions, whereas new interactions may be required to allow transitions from one step to the next. A possible outcome of these structural transitions could be: (1) interconversion of the secondary structures of the RNA components of the splicing machinery from inactive to active forms and vice versa; (2) dissociation of RNA components of the splicing machinery so they can leave or become accessible to other factors; or (3) formation of new structures for transition into new steps. Similar models have been suggested for the function of MSS116, a putative RNA helicase involved in yeast mitochondrial mRNA splicing⁵⁷. Alternatively, some of these PRP proteins may have a role in a proofreading mechanism as has been suggested for PRP16 protein³⁶. In respect of the role of PRP22 in pre-mRNA splicing processes, it is tempting to speculate that the unwinding of a specific RNA-RNA interaction by PRP22 is required for the release of the mRNA from the spliceosome.

How do these putative helicases achieve their specificity? The five PRP proteins discussed above have variable domains in addition to the helicase-like domain (Fig. 6b). The variable domain might dictate the specificity for the helicase function through interaction with either the protein or RNA components of the splicing machinery. In this regard the S1 RNA-binding motif in the N-terminal region of PRP22 may mediate the specific interaction of PRP22 with its substrate. □

Received 21 November; accepted 11 December 1990.

- Brody, E. & Abelson, J. *Science* **228**, 963-967 (1985).
- Freundewey, D. & Keller, W. *Cell* **42**, 355-367 (1985).
- Grabowski, P. J., Seiler, S. R. & Sharp, P. A. *Cell* **42**, 345-353 (1985).
- Green, M. R. A. *Rev. Genet.* **20**, 671-708 (1986).
- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. *Rev. Biochem.* **55**, 1119-1150 (1986).
- Cech, T. R. & Bass, B. L. *Rev. Biochem.* **55**, 599-629 (1986).
- Sharp, P. A. *Science* **235**, 766-771 (1987).
- Pikielny, C. W., Rymond, B. C. & Rosbash, M. *Nature* **324**, 341-345 (1986).
- Cheng, S. C. & Abelson, J. *Genes Dev.* **1**, 1014-1027 (1987).
- Konarska, M. M. & Sharp, P. A. *Cell* **49**, 763-774 (1987).
- Lin, R. J., Lustig, A. J. & Abelson, J. *Genes Dev.* **1**, 7-18 (1987).
- Sawa, H., Ohno, M., Sakamoto, H. & Shimura, Y. *Nucleic Acids Res.* **16**, 3157-3164 (1988).
- Vijayraghavan, U. & Abelson, J. *Molec. cell. Biol.* **10**, 324-332 (1990).
- Schwer, B. & Guthrie, C. *Nature* **349**, 494-499 (1991).
- Maniatis, T. & Reed, R. *Nature* **325**, 673-678 (1987).
- Guthrie, C. & Patterson, B. A. *Rev. Genet.* **22**, 387-419 (1988).
- Steitz, J. A. *et al.* in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles* (ed. Birnstiel, M. L.) 115-154 (Springer-Verlag, Berlin, 1988).
- Lührman, R. in *Structure and Function of Major and Minor Small Nuclear ribonucleoprotein Particles* (ed. Birnstiel, M. L.) 71-99 (Springer-Verlag, Berlin, 1988).
- Vijayraghavan, U. & Abelson, J. *Nucleic Acids molec. Biol.* **3**, 197-215 (1989).
- Woolford, J. L. *Yeast* **5**, 439-458 (1989).
- Ruby, S. W. & Abelson, J. *Trends Genet.* **7**, 1-9 (1991).
- Lustig, A. J., Lin, R. J. & Abelson, J. *Cell* **47**, 953-963 (1986).
- Lossky, M. G., Anderson, J., Jackson, S. P. & Beggs, J. D. *Cell* **51**, 1019-1026 (1987).
- Pinto, A. L. & Steitz, J. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8742-8746 (1989).
- Anderson, G. J., Bach, M., Lührmann, R. & Beggs, J. D. *Nature* **342**, 819-821 (1989).
- Banroques, J. & Abelson, J. *Molec. cell. Biol.* **9**, 3710-3719 (1989).
- Bjorn, S. P., Soltyk, A., Beggs, J. D. & Friesen, J. D. *Molec. cell. Biol.* **9**, 3698-3709 (1989).
- Chang, T. H., Clark, M. W., Lustig, A. J., Cusick, M. E. & Abelson, J. *Molec. cell. Biol.* **8**, 2379-2393 (1988).
- Hodgman, T. C. *Nature* **333**, 22-23 (1988).
- Lane, D. *Nature* **334**, 478 (1988).
- Gorbalenya, E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. *Nucleic Acids Res.* **17**, 4713-4730 (1989).
- Gorbalenya, A. E. & Koonin, E. V. *Nucleic Acids Res.* **17**, 8413-8440 (1989).
- Linder, P. *et al.* *Nature* **337**, 121-122 (1989).
- Chang, T. H., Arenas, J. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1571-1575 (1990).

- Dalbadie-McFarland, G. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4236-4240 (1990).
- Burgess, S., Couto, J. R. & Guthrie, C. *Cell* **60**, 705-717 (1990).
- Chen, J. H. & Lin, R. J. *Nucleic Acids Res.* **18**, 6447 (1990).
- Vijayraghavan, U., Company, M. & Abelson, J. *Genes Dev.* **3**, 1206-1216 (1989).
- Lamond, A. I., Konarska, M. M. & Sharp, P. A. *Genes Dev.* **1**, 532-543 (1987).
- Sikorski, R. S. & Heiter, P. *Genetics* **122**, 19-27 (1989).
- Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
- Schnier, J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 1008-1011 (1982).
- Subramanian, A. R. *Prog. nucleic Acids. Res. molec. Biol.* **28**, 101-142 (1983).
- Regnier, P., Grunberg-Manago, M. & Portier, C. *J. biol. Chem.* **262**, 63-68 (1987).
- Subramanian, A. R. *Essays biochem.* **21**, 45-85 (1985).
- Gribskov, M., Lüthy, R. & Eisenberg, D. *Meth. Enzymol.* **113**, 146-159 (1989).
- Sands, J. F., Cummings, H. S., Sacerdot, C., Dondon, L., Grunberg-Manago, M. & Hershey, J. W. B. *Nucleic Acids Res.* **15**, 5157-5168 (1987).
- Couto, J. R., Tamm, J., Parker, R. & Guthrie, C. *Genes Dev.* **1**, 445-455 (1987).
- Lain, S., Riechmann, J. L., Martin, M. T. & Garcia, J. A. *Gene* **82**, 357-362 (1989).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945-951 (1982).
- Fry, D. C., Kuby, S. A. & Mildvan, A. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 907-911 (1986).
- Ray, B. K. *et al.* *J. biol. Chem.* **260**, 7651-7658 (1985).
- Rozen, F. *et al.* *Molec. cell. Biol.* **10**, 1134-1144 (1990).
- Hirling, H., Scheffner, M., Restle, T. & Stahl, H. *Nature* **339**, 562-564 (1989).
- Lain, S., Riechmann, J. L. & Garcia, J. A. *Nucleic Acids Res.* **18**, 7003-7006 (1990).
- Lain, S., Martin, M. T., Riechmann, J. L. & Garcia, J. A. *J. Virol.* **65**, 1-6 (1991).
- Seraphin, B., Simon, M., Boulet, A. & Faye, G. *Nature* **337**, 84-87 (1989).
- Schnier, J. *et al.* *Nucleic Acids Res.* **16**, 3075-3089 (1988).
- Schnier, J. & Faist, G. *Molec. gen. Genet.* **200**, 476-481 (1985).
- Umesono, K., Inokuchi, H., Ohyama, K. & Ozeki, H. *Nucleic Acids Res.* **12**, 9551-9565 (1984).
- Sijben-Mueller, G., Halliok, R. B., Alt, J., Westhoff, P. & Herrmann, R. G. *Nucleic Acids Res.* **14**, 1029-1044 (1986).
- Cigan, A. M., Pabich, E. K., Feng, L. & Donahue, T. F. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2784-2788 (1989).
- Ernst, H., Duncan, R. F. & Hershey, J. W. B. *J. biol. Chem.* **262**, 1206-1212 (1987).
- Lain, S., Riechmann, J. L. & Garcia, J. A. *Virus Res.* **13**, 157-172 (1989).
- Collett, M. S. *et al.* *Virology* **165**, 191-199 (1988).
- Madkow, E. *et al.* *Virology* **159**, 217-228 (1987).
- Nielsen, P. J., McMaster, G. K. & Trachsel, H. *Nucleic Acids Res.* **13**, 6867-6880 (1985).

ACKNOWLEDGEMENTS. We thank M. Gribskov for his assistance on the use of PROFILESEARCH Program, and T.-H. Chang and C. O'Day for comments on the manuscript. This work was supported by the NIH (J.A.).

PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome

Beate Schwer & Christine Guthrie

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

The assembly of the spliceosome is an ATP-dependent process. The splicing factor PRP16 contains variations of several motifs that define the eIF-4A-like ATP-dependent RNA helicase family. The protein has now been purified and shown to exhibit RNA-dependent ATPase activity. PRP16 is required specifically for the second catalytic step of the splicing reaction *in vitro*. This function requires ATP binding and/or hydrolysis, which appears to be concomitant with release of the protein from the spliceosome. PRP16 may be the prototype for a set of splicing factors which use ATP to drive a cycle of conformational changes.

SPlicing of pre-mRNA occurs in a large and structurally dynamic complex, the spliceosome. Within this structure, two distinct transesterification reactions take place which are mechanistically similar to those of self-splicing by group II introns¹⁻³. The first step involves cleavage at the 5' splice-site and formation of a 2'-5' branched intermediate (lariat); in the second step the 3' splice-site is cleaved and the exons are ligated. However, in contrast to the highly conserved structural elements that reside within group II introns⁴, the only conserved features of pre-mRNA introns lie within short regions at or near the splice junctions^{5,6}. Many *trans*-acting factors are involved in pre-mRNA splicing, presumably compensating for the low information content in the intron. Four ribonucleoprotein particles composed of small nuclear RNA (snRNA) (U1, U2, U4/U6 and U5) and proteins (snRNPs)^{7,8}, together with an as yet unknown number of non-snRNP proteins, interact with the pre-mRNA to assemble the spliceosome in a highly ordered, ATP-dependent pathway^{9,10}. Because the two-step transesterification mechanism of self-splicing is energy-independent, a major unsolved question is the role of ATP in the assembly of the spliceosome. This problem could be addressed by the identification and characterization of spliceosomal ATPases.

The biochemical analysis of the splicing process in yeast has been facilitated by the use of genetics. In particular, many essential proteins have been uncovered by conditional lethal mutations; the pre-mRNA processing (*PRP*) mutants are defective in splicing at the non-permissive temperature *in vivo*^{11,12}, and in many cases, exhibit temperature-sensitive activity *in vitro*^{13,14}. Several of the PRP proteins have been shown to be tightly associated with specific snRNAs and/or splicing complexes. For example, PRP8 (ref. 15) and PRP4 (refs 16, 17) are found in association with U5 and U4/U6, respectively, and PRP11 (ref. 18) and PRP8 (ref. 19) are integral components of the 40S spliceosome. PRP2, in contrast, appears to be an extrinsic splicing factor required in a late stage of spliceosome formation, prior to cleavage at the 5' splice site²⁰. As yet, no specific biochemical activity has been assigned to any of the proteins.

Another genetic approach to the identification of splicing factors is the isolation of second-site suppressors. A potential advantage of this strategy is that the mutant phenotype may be

indicative of functional interactions between spliceosomal components. The *prp16-1* allele was initially isolated as a dominant suppressor of a branchpoint mutation (A to C) which causes the accumulation of unspliced pre-mRNA²¹. Cells containing *prp16-1* use the mutant branchpoint nucleotide with increased efficiency. In addition, the *prp16-1* mutation has a recessive phenotype on the splicing of wild-type messages, and both pre-mRNA and lariat intermediates accumulate^{21,22}. This complex phenotype could reflect a role for PRP16 in both steps of the splicing reaction. Intriguingly, the sequence of the *PRP16* gene product contains homologies to motifs that define a family of proteins whose prototype, eIF-4A, is known to be an ATP-dependent RNA helicase²³. Within most of these motifs, however, the sequence of *PRP16* exhibits variations from the helicase consensus; for example, the sequence of the 'DEAD box'²⁴ is, in *PRP16*, DEAH²². Notably, the mutation which confers the *prp16-1* suppressor phenotype was shown to reside within the putative ATP-binding domain, suggesting that the accuracy of branchpoint recognition is coupled to the binding and/or hydrolysis of ATP²².

We have now used a cell-free system to investigate the function of PRP16. We report here the purification of this protein and show that PRP16 is required, with ATP, for the second cleavage-ligation reaction *in vitro*. We also show that PRP16 is a non-snRNP protein that associates tightly with spliceosomes that have completed 5' cleavage and lariat formation. This interaction can be detected only in the absence of ATP; upon addition of ATP, PRP16 appears to be released from the spliceosome. Finally we show that purified PRP16 exhibits RNA-dependent ATPase activity. The recent sequence analysis of two other *PRP* genes, each of which contains the DEAH-type variant motif, suggests that PRP16 may be the prototype for a set of splicing factors that use a common mechanistic strategy to promote distinct steps in the pathway.

PRP16 in *in vitro* splicing reaction

Antibodies were raised against a trpE-PRP16 fusion protein purified from *E. coli* (see methods, Fig. 1). The anti-PRP16 antiserum, but not pre-immune serum, recognized a protein of the predicted size²² of relative molecular weight 120,000 on a western blot (Fig. 1a). To confirm that the antibodies recognize the authentic PRP16 protein, we expressed the coding region of PRP16 on a 2 μ plasmid under the control of the constitutive glucose-6-phosphate dehydrogenase promoter²⁵. In Fig. 1a, equal amounts of protein were loaded in each lane; as expected, the levels of PRP16 were significantly increased in extracts from the overproducing strain.

The effect of anti-PRP16 antibodies in an *in vitro* splicing reaction is shown in Fig. 1b. Extract was pre-incubated with protein A-purified anti-PRP16 antiserum (1–10 μ g IgG) and then assayed for splicing of a labelled actin precursor for 20 minutes at 24 °C²⁶. The first step of the splicing reaction, 5' cleavage and lariat formation, occurred efficiently. However, no mature RNA was formed and splicing intermediates accumulated. To distinguish whether PRP16 is actually dispensable for spliceosome assembly and the first cleavage-ligation reaction, or whether the antibodies selectively inhibit the second step, we immuno-depleted yeast splicing-extract for PRP16 (referred to as Δ PRP16). Depletion was analysed by western blot analysis;

no remaining PRP16 could be detected (detection limit was 5–10%, data not shown). As shown in Fig. 2, Δ PRP16 carried out the first step of the splicing reaction and accumulated intermediates.

Further evidence that PRP16 is required only for the second catalytic reaction *in vitro* comes from analysis of a temperature-sensitive allele of *PRP16*. The splicing mutant *prp23*, isolated by Vijayraghavan *et al.*¹², is allelic to *PRP16* (data not shown). When *prp23* cells are grown at the restrictive temperature, precursor and lariat intermediates accumulate (ref. 12, and unpub-

lished results); heat-inactivation of *prp23* extract results in the accumulation of splicing intermediates and the first step of the reaction is unaffected (data not shown). Thus, unlike other splicing proteins such as PRP11 (ref. 18) or the U5-specific protein PRP8 (ref. 19) which have been shown to be part of the 'core'-spliceosome, PRP16 is dispensable for the splicing reaction *in vitro* until the second step. To investigate the possible association of PRP16 with snRNAs, we performed immunoprecipitation experiments using anti-PRP16 antibodies. We found that antibodies against the trimethylguanosine cap structure of snRNAs (anti-TMG) but not anti-PRP16 antiserum could immunoprecipitate snRNAs (as monitored by ³²P-pCp labelling using RNA ligase²⁷) from yeast splicing extracts (data not shown). The finding that PRP16 is not tightly associated with any snRNA is consistent with the data that it is not required for step 1.

Complementation analysis

We sought to complement the second-step defect in PRP16-immuno-depleted extracts by the addition of fractions from

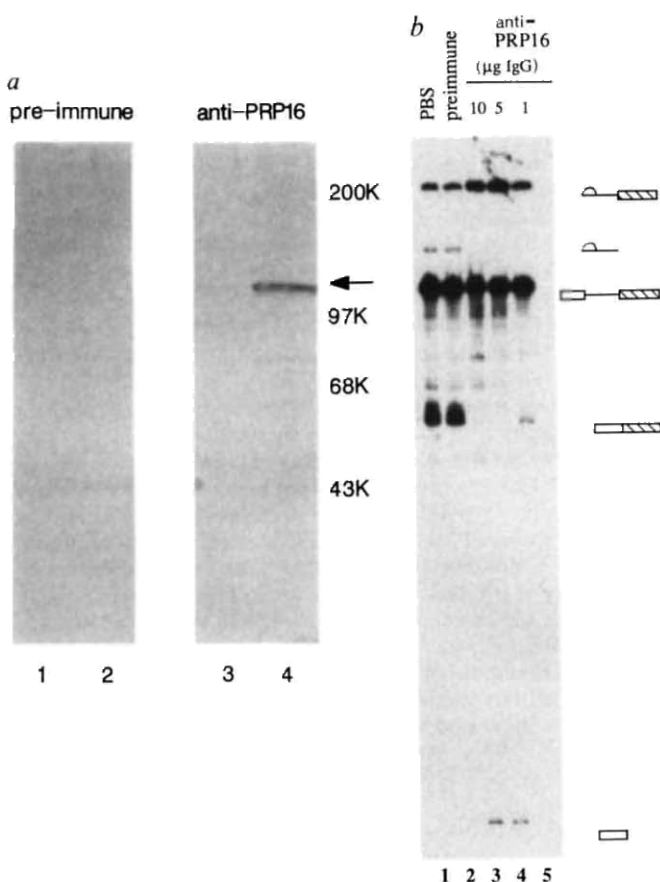


FIG. 1 a, Western blot. Proteins in yeast extracts were separated by SDS-PAGE and transferred to nitrocellulose⁴³. The filter was incubated with anti-PRP16 (1:1,500 dilution) (lanes 3, 4) or pre-immune serum (lanes 1, 2) followed by goat anti-rabbit secondary antibodies conjugated to alkaline phosphatase (BIO-RAD). Yeast extracts were prepared from YTIM1 (a/ α , *lys2/lys2*, *trp1/trp1*, *his3/his3*, *ura3/ura3*, *LEU2/leu2*)²² (lanes 1, 3) and YTIM1/pG1PRP16 (lanes 2, 4). The positions of migration of pre-stained molecular weight markers (BRL) are indicated. b, Inhibition of an *in vitro* splicing reaction by anti-PRP16. 4 μ l of splicing extract, prepared as previously described²⁶, was pre-incubated for 10 min on ice with 1 μ l of PBS (lane 1), pre-immune serum (lane 2), anti-PRP16 (10, 5, 1 μ g IgG) (lanes 3, 4, 5) and then incubated under standard splicing conditions²⁶ in a final volume of 10 μ l for 25 min at 24 °C. The reactions were analyzed by denaturing PAGE. The precursor was a [α -³²P]GTP-labelled T7 transcript of actin²⁶ synthesized in the presence of 500 μ M GpppG (Pharmacia) (the plasmid was provided by Art Zaig). The precursors, reaction intermediates and products are represented schematically on the right.

METHODS. Antibodies were prepared using a fusion protein containing residues 40–698 of the *prp16-1* protein fused to *E. coli* TrpE expressed from plasmid pATH2 (ref. 44). The hybrid protein was induced with indoleacrylic acid, purified as previously described⁴⁵, and used to immunize rabbits (BAbCO, Inc., Berkeley, California). Polyclonal antiserum was purified using protein A-Sepharose (Pharmacia). Overexpression of PRP16 on plasmid pG1PRP16: a *NDE* I site has been generated at the start codon of PRP16 (position +1)²² by site-directed mutagenesis. Then the *Nde*I-*Sph*I fragment (nucleotides –1 to +3275) of PRP16 was treated with T4 polymerase and ligated into the pG1 (ref. 25) *Sal*I (T4-polymerase treated) site. pG1PRP16 was transformed into YTIM1 using the lithium acetate method⁴⁶.

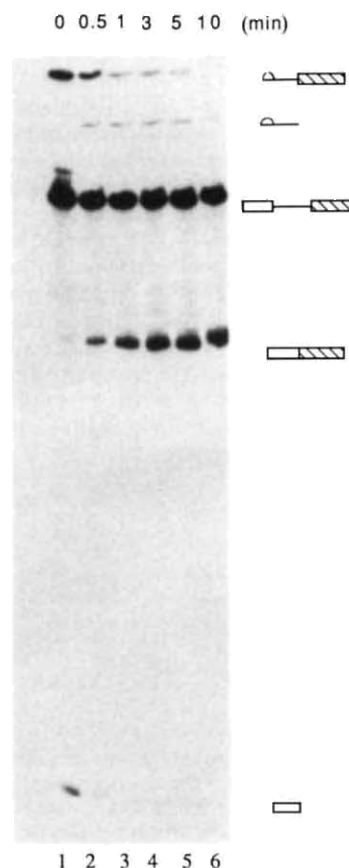


FIG. 2 Complementation of the second-step block in Δ PRP16 extracts. Splicing extract was immuno-depleted for PRP16 and then incubated with ³²P-labelled actin precursor for 20 min at 24 °C. After this incubation, fraction I (prepared from YTIM1 according to ref. 28) was added and aliquots analysed after 0–10 min on a denaturing polyacrylamide gel.

METHODS. For the immuno-depletion, 60 μ l of splicing extract was incubated with 5 μ l of protein A-purified anti-PRP16 antiserum (25 μ g IgG per μ l) for 1 h on ice. 60 μ l protein A-Sepharose (Pharmacia) (1 g protein A-Sepharose per 10ml) was washed in PBS and briefly centrifuged. The extract, which had been pre-incubated with the antiserum, was added to the beads with the addition of 25 μ l of PBS. After incubation for 1 hour at 4 °C on a nutator, the beads were removed by centrifugation and the supernatant assayed for splicing. 6 μ l was used per splicing reaction in a final volume of 10 μ l. The final concentrations were: 2mM ATP, 2.5 mM MgCl₂, 20 mM KCl, 60 mM potassium phosphate, 8 mM HEPES, 3% PEG 8000, 80 μ M EDTA, 0.5 mM DTT, 8% glycerol, 2 mM sodium phosphate, 30 mM NaCl, 0.5 fmol precursor.

wild-type yeast extracts. Western blot analysis indicated that PRP16 is enriched in fraction I (40P3), a 40% ammonium sulphate precipitate of yeast splicing extracts²⁸ (data not shown). Δ PRP16 extract was incubated under splicing conditions for 25 min at 24 °C during which time splicing intermediates accumulated (Fig. 2). 40P3 was then added and aliquots analysed after different times as indicated in Fig. 2. The formation of mature RNA could be detected as early as 30 s after the addition of 40P3. As the lag-phase for the production of mRNA in a splicing reaction is at least 5 min (ref. 26, and data not shown), the rapid kinetics for complementation in Δ PRP16 extracts argues against the possibility that spliceosomes dissociate and then reassemble upon the addition of PRP16. Instead, the rapid appearance of mature RNA and the concomitant decrease in intermediates suggest that PRP16 associates with the pre-assembled spliceosomes to promote the second step. The complementing activity is resistant to micrococcal nuclease treatment (up to 15,000 U per μ l, 10 min, 30 °C) and sensitive to heat treatment (60 °C, 5 min) (data not shown).

'Chase' experiments

To test the notion that the spliceosomes pre-formed in Δ PRP16 extracts are functional intermediates in the complementation assay, the reaction was challenged at various times by the addition of unlabelled pre-mRNA (Fig. 3). When an excess of ~500-fold unlabelled pre-mRNA was added together with ³²P-labelled precursor to an *in vitro* splicing reaction, it competed with the labelled precursor for splicing and no mRNA or intermediates in Δ PRP16 extracts were formed. But if splicing has already proceeded through the first step in Δ PRP16 extracts in the presence of only labelled pre-mRNA, labelled intermediates accumulated. At this point, fraction I was added either alone (lanes 5–7) or together with unlabelled precursor (lanes 8–10). If spliceosomes first dissociated and then reassembled with factors present in the complementing fraction, one would not expect to see products in the presence of unlabelled precursor. Products formed to the same extent and with the same kinetics in both reactions (lanes 5–7 and 8–10 respectively), demonstrat-

ing unambiguously that PRP16 can associate with the pre-formed spliceosome to function.

Purification of PRP16

The PRP16 protein was purified from a yeast strain YTdL4/E3 containing the PRP16 coding-region on a high copy-number plasmid under the transcriptional control of the highly expressed constitutive GPD promoter²⁵ (Fig. 1a) which results in ~100-fold overexpression. The level of overproduction was assayed by functional complementation of the second step block in Δ PRP16 extracts; fraction I from the overproducing strain is about 100 times more active than fraction I from a wild-type strain. This was also confirmed by quantitative western blot analysis using ¹²⁵I-labelled secondary antibodies (data not shown). An epitope of 10 amino acids had been linked to the carboxy-terminus of the protein (Fig. 4). This peptide does not interfere with the function of PRP16 either *in vivo* or *in vitro*: one copy of the tagged gene can support growth of a strain containing a disruption of the *PRP16* locus and fraction I prepared from this strain can complement the second-step splicing defect in Δ PRP16 extracts (data not shown). The epitope tag provides the possibility of detecting the protein using a well characterized monoclonal antibody, KT3 (ref. 29), specific to eight of the ten added amino acids. From this strain, we purified PRP16 to at least 90% homogeneity by ammonium-sulphate precipitation, phosphocellulose and CM-Sephacryl chromatography (Fig. 4a). We assayed the purification by functional complementation of the second-step defect in heat-inactivated *prp23* extracts. Because the gel assay for splicing is not easily amenable to quantitation, we could not precisely calculate the overall purification of PRP16. However, when the protein was added in limiting amounts in the complementation assay, the amount of spliced mRNA generated was approximately proportional to the concentration of PRP16 and to the incubation time. We defined 1 unit as the activity that resulted in a half-maximal amount of mRNA produced in 4 min when added to pre-formed intermediates. The activity in the starting extract could not be measured as it is splicing competent by itself, and

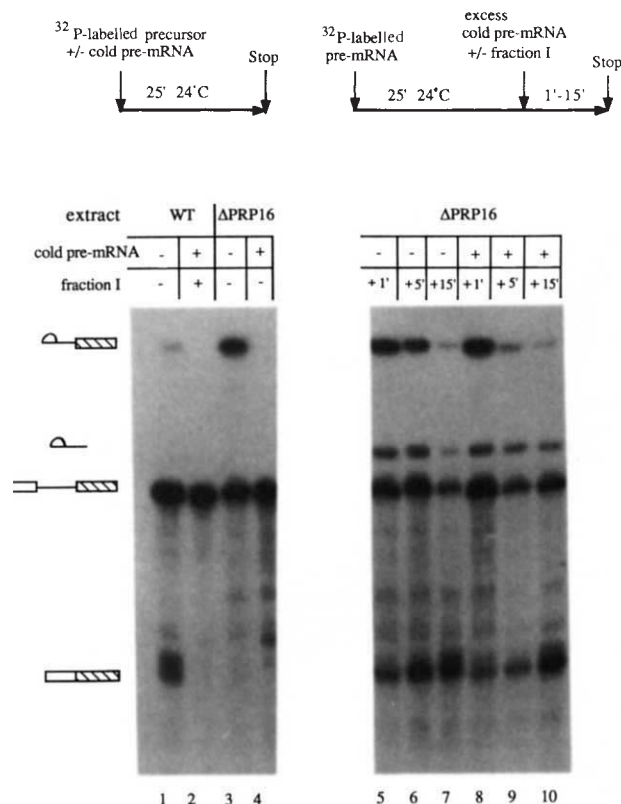


FIG. 3 Splicing intermediates are 'chased' into products. Standard splicing reactions were carried out in the absence (lane 1) or in the presence (lane 2) of a 500-fold excess of precursor RNA. Extract has been immunodepleted for PRP16 (lanes 3–10). Δ PRP16 extract was incubated with ³²P-labelled actin precursor either without (lane 3) or with (lane 4) the addition of a 500-fold excess of unlabelled pre-mRNA. Lanes 5–10, Δ PRP16 extract was incubated under standard splicing conditions²⁶ with ³²P-labelled precursor for 25 min at 24 °C. Fraction I was then added either alone (lanes 5–7) or together (lanes 8–10) with an excess of unlabelled precursor and the incubation continued for 1–15 min. The amount of RNA loaded in lane 9 is lower than in the other lanes.

was arbitrarily set at 100% (Fig. 4b). With this assumption, the overall purification was about 400-fold from the overproducing strain and the recovery was approximately 87% (Fig. 4b).

RNA-dependent ATPase activity

The PRP16 protein contains motifs homologous to sequences defining a family of RNA-dependent ATPases²². The first step to investigate whether PRP16 requires ATP for its function was to test whether the purified protein exhibits ATPase activity. ATP hydrolysis was measured as the release of $^{32}\text{P}_i$ from ^{32}P - γATP using Norit-charcoal³⁰. PRP16 exhibited ATPase activity which was stimulated substantially by the presence of RNA (Fig. 5). Poly(A) was the most effective, increasing ATPase activity by a factor of 8–9-fold, whereas single-stranded M13 DNA had no detectable effect. The specific activity of this PRP16 preparation was 54 pmol P_i released per μg per min, which is one to two orders of magnitude higher than that reported for eIF-4A (ref. 31), three times higher than srmB (ref. 32) and about one order of magnitude less than p68 (ref. 33). Thus the RNA-dependent ATPase activity of PRP16 is within the range of the activities measured for other members of this family. Both ATPase activity and the activity complementing the splicing defect of heat-inactivated *prp23* extracts cosediment with the PRP16 protein in a sucrose gradient (data not shown). To confirm that the ATPase activity is due to PRP16 and not to a contaminating protein, we made use of the fact that PRP16 was epitope-tagged. The peptide at the carboxy terminus is specifically recognized by the monoclonal antibody KT3 (ref. 29). As the antibody-binding epitope lies outside the ATP-binding domain of PRP16, it was conceivable that the immunoprecipitated protein would retain ATPase activity even

when bound to protein G-Sepharose beads. This approach has been successfully used to measure the RNA-dependent ATPase activity of p68 (ref. 34). Immunoprecipitations were performed using the purified tagged protein and the RNA-dependent ATPase activity was assayed in the immunoprecipitate. The efficiency of immunoprecipitation, as assayed by western blotting, was 15–20% of tagged protein; correspondingly, the measured RNA-dependent ATPase activity in the pellet was ~20% of the input activity (data not shown).

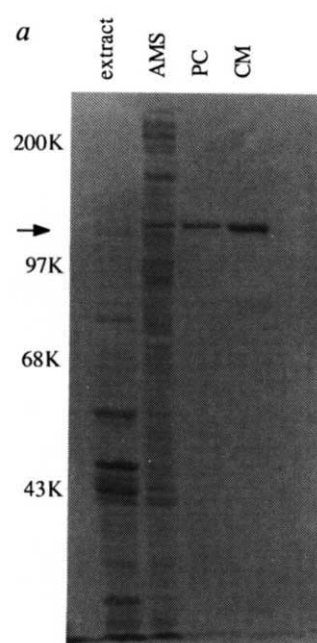
PRP16 binds to the spliceosome

As purified PRP16 shows ATPase activity, it is an important question to test if the rapid conversion of splicing intermediates into mRNA, which is dependent on PRP16, is correlated with an ATP requirement. The scheme of this experiment is outlined in Fig. 6. Pre-mRNA was incubated in ΔPRP16 extracts for 20 min to allow accumulation of lariat intermediates and exon I (Fig. 6). ATP was then removed by gel-filtration. As shown in Fig. 6, the addition of buffer, ATP or PRP16 did not result in the formation of mRNA, but when ATP and PRP16 were added together, the reaction proceeded through the second step. This result confirms the previously described requirement for ATP in the second cleavage-ligation reaction^{14,35} and is likely to reflect the ATPase activity of PRP16.

We have shown that PRP16 can associate with pre-formed spliceosomes to promote step two of the splicing reaction. In initial experiments, however, we were unsuccessful in detecting the protein in the spliceosome, as assayed by immunoprecipitation of splicing products by anti-PRP16 antibodies. We therefore chose the strategy outlined in Fig. 6; the rationale was that PRP16 might bind transiently to the spliceosome but be released

FIG. 4 Purification of PRP16. *a*, Proteins in crude splicing extract (extract) and pooled active fractions purified by 40% ammonium sulphate precipitation (AMS), phosphocellulose (PC) and carboxymethyl-Sepharose (CM) were separated on a 7.5% SDS-PAGE gel and silver stained⁴⁷. Lanes 1, 2: 10 μg protein; lanes 3, 4: 1 μg protein. The protein concentrations were measured according to Bradford⁴⁸. Prestained high-molecular-weight markers (BRL) were used as standards. *b*, Purification table.

METHODS. YTDL4: nucleotides +119 to +3275 of PRP16²² have been replaced by a 4.8 kb *EcoRI*-*Clal* fragment of *LYS2* (ref. 49) in a haploid strain derived from YTIM1 (ref. 22). YTDL4/E3: YTDL4 strain containing the tagged PRP16 on the pG1 plasmid. A *HindIII* site was introduced at the C-terminus of PRP16 by site-directed mutagenesis; a 120-nucleotide PCR fragment containing the T antigen carboxy terminus was ligated into this site resulting in the addition of the peptide sequence KLTPPEPET at the 3' end of PRP16. The peptide TPPPEPET is recognized by the monoclonal antibodies KT3 (ref. 29). For the purification, 12 l of YTDL4/E3 were grown in YEPD (to 5×10^7 cells ml^{-1}) and splicing extract prepared (105 ml) according to ref. 26 except that lyticase (Sigma) was used instead of zymolyase. The 40% ammonium sulphate precipitation was performed according to the protocol used to prepare fraction I (ref. 28). 12 ml of fraction I was diluted fourfold in buffer D (20 mM HEPES, pH 7, 20% glycerol, 50 mM KCl, 0.2 mM EDTA, 0.5 mM DTT) and purified over a 50 ml phosphocellulose (Whatman) column, washed with 5 column-volumes of buffer D-250 (buffer D containing 250 mM KCl), and eluted in a linear 250 mM–600 mM KCl gradient (6 column-volumes). The active pooled fractions (50 ml) were dialysed for 3.5 h against buffer M (20 mM MES, pH 6.6, 50 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 10% glycerol) and applied to a 10-ml CM-Sepharose (Pharmacia) column equilibrated in buffer M, washed with 5 column-volumes of the same buffer and eluted with a linear 50–300 mM KCl gradient in buffer M (10 column-volumes). The buffers usually contained 0.5 nM PMSF, 1 μg ml^{-1} leupeptin and 0.5 μg ml^{-1} pepstatin. The activity was assayed as follows: *prp23* extract was heat-inactivated at 37 °C for 20 min and then incubated with precursor under splicing conditions²⁶ for 20 min at 24 °C. To 10 μl of this mix, 1 μl of diluted (1:40 for PC and 1:400 for CM fractions, respectively, in buffer D) column fractions was added and the incubation continued for 4 min at 24 °C. The splicing reactions were stopped, phenol-extracted twice and analysed by denaturing PAGE. To determine the activity, serial dilutions of the pooled fractions were analysed for complementation. The maximum activity was set by the amount of mRNA (in c.p.m.) produced by 1 μl of fraction I in 4 min. One unit was defined as the amount of the fraction resulting in half-maximal amount of mRNA in 4 min.



b

Fraction	Protein (mg)	Yield (%)	Specific activity [units per mg]	Cumulative purification
Splicing extract	928	100*	ND	×1
40% ammonium sulphate	207	100*	3.9×10^4	×4.5
Phosphocellulose	7.7	85	9×10^5	×102
CM-Sepharose	2	87.5	3.5×10^6	×406

* As the activity could not be assayed in the crude extract (ND), this value was set arbitrarily at 100%.

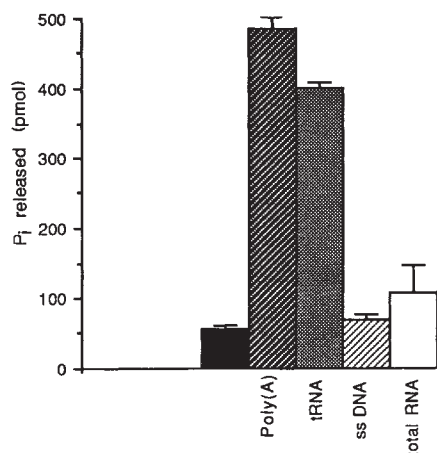


FIG. 5 RNA-dependent ATPase activity. ATP-hydrolysis was measured in the absence (■) and presence of 250 $\mu\text{g ml}^{-1}$ RNA/DNA (as indicated). Stimulation by poly(A) 8.5 $\times$, by poly(U) 4.3 $\times$, by poly(C) 2.6 $\times$, by tRNA 7.1 $\times$, by single-stranded M13 DNA (ssDNA) 1.2 $\times$, by ribosomal RNA 1.8 $\times$, and by total RNA 1.9 $\times$.

METHODS. 400 ng PRP16 (2 μl of the CM fraction) was incubated in 50 mM triethanolamine, pH 8, 75 mM potassium acetate, 1 mM DTT, 0.5 mM MgCl_2 , 100 μM ATP, [γ - ^{32}P]ATP (1 μCi , 10 Ci mmol^{-1}) without or with the addition of 10 μg of RNA or single-stranded M13 DNA. After incubation for 20 min at 24 $^{\circ}\text{C}$, unreacted ATP was precipitated by the addition of 260 μl of acid-washed Norit-charcoal (5% in 20 mM phosphoric acid)³⁰. After centrifugation (10 min), 100 μl of the supernatant was analysed for free $^{32}\text{P}_i$ by Cerenkov counting. The background of about 40 pmol resulting from incubation in the absence of PRP16 was subtracted from each value. The results are average values of at least duplicate reactions, the bars show the standard deviation. Stimulation is given as the fold increase over incubation in the absence of any nucleic acids.

upon hydrolysis of ATP. The reactions were tested for immunoprecipitability by anti-PRP16 antiserum (Fig. 6). Interestingly, we found that PRP16 could bind to the spliceosome in the absence of ATP, resulting in the immunoprecipitation of splicing intermediates (Fig. 6). In contrast, spliceosomes were not immunoprecipitated in the presence of ATP and PRP16, suggesting that the interaction is destabilized or structurally altered (making the epitopes inaccessible to antibodies) upon the addition of ATP.

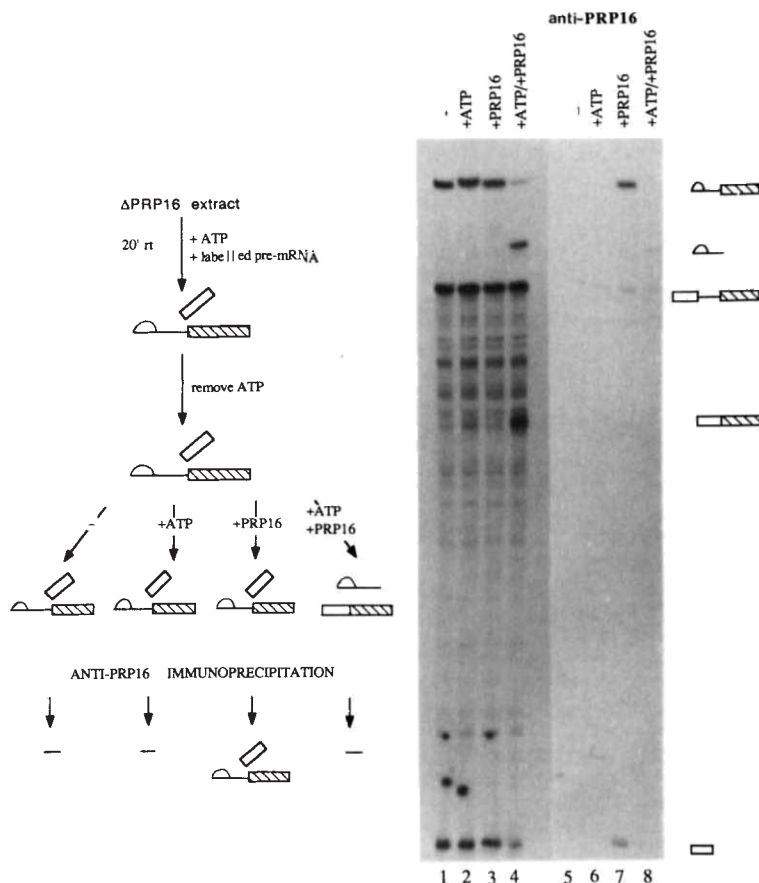
Discussion

Using PRP16-immunodepleted extract (ΔPRP16) we have shown that the first catalytic step of the splicing reaction can occur *in vitro* in the absence of PRP16, but that the protein is absolutely required for the second cleavage-ligation reaction.

Based on functional complementation of the second step block, we have purified PRP16 from yeast. By the immunoprecipitation of lariat-intermediates with anti-PRP16 antibodies, we have demonstrated that PRP16, a non-snRNP protein, can bind tightly to spliceosomes that have completed the first cleavage-ligation reaction. Moreover, promotion of the second step by PRP16 is ATP-dependent, since ΔPRP16 extracts depleted of ATP by gel-filtration require the addition of ATP and PRP16 to generate ligated exons. But splicing intermediates (or products) cannot be immunoprecipitated with anti-PRP16 antibodies if ATP is present. This finding suggests that ATP binding and/or hydrolysis promotes the destabilization or release of PRP16 from the spliceosome. Alternatively, the protein may remain bound, but undergo a conformational rearrangement that masks the antigenic epitopes. In any case, PRP16 transiently associates

FIG. 6 The second step of splicing requires PRP16 plus ATP. Splicing extract was immuno-depleted of PRP16 and incubated under splicing conditions for 20 min, then the reaction mixture was loaded onto a G50-Sephadex spin-column equilibrated in splicing buffer without precursor and ATP. 40 μl of the flowthrough was used and the incubation continued for 10 min after the addition of either 1 μl buffer (—), 1 μl 80 mM ATP (+ATP), 1 μl CM fraction (+PRP16) or both (+ATP/+PRP16). 10 μl was analysed (lanes 1–4) and 30 μl of each reaction used for immunoprecipitation experiments (lanes 5–8).

METHODS. Immunoprecipitation: 5 μl of anti-PRP16 antiserum was pre-incubated with protein A-Sepharose beads (40 μl of 1 g protein A-Sepharose per 10 ml) in 400 μl IPP₅₀₀ (ref. 50) for 1 h at 24 $^{\circ}\text{C}$ and then washed extensively (3 times in IPP₁₅₀). 200 μl of IPP₁₅₀ and 30 μl of the splicing reactions were added to the beads and incubated for 1 h at 4 $^{\circ}\text{C}$ on the nutator. The beads were then washed three times with 400 μl of IPP₁₅₀, phenol-extracted and bound RNA was analysed by denaturing PAGE.



with the spliceosome, as our experiments show that functional spliceosomes can form efficiently in the absence of the protein and can be rapidly chased into products upon addition of PRP16.

We have also shown that highly purified PRP16 has ATPase activity, which is significantly stimulated by RNA but not by single-stranded DNA. This result confirms that PRP16 is a member of the RNA-dependent ATPase family proposed by Burgess *et al.*²² In conjunction with the above results, the simplest interpretation is that the PRP16 protein itself mediates the previously observed ATP requirement for the second cleavage-ligation reaction^{14,35}. Whether PRP16 is the only ATPase required for the second step must await further analysis. Similarly, future experiments are required to determine whether the protein also plays a part in earlier steps of the pathway, as the genetic data argue. The *prp16-1* allele was isolated by its ability to alleviate the first-step splicing defect conferred by a branchpoint mutation²¹, and the suppressor mutation was shown to lie in the proposed ATP-binding domain of the protein²². Thus, PRP16 may have an additional (ATP-dependent) role which is not measurable in the cell-free system. Discrepancies between *in vivo* and *in vitro* requirements have also been observed in the study of transcription reactions³⁶.

Understanding the role of ATP in the function of PRP16 is particularly interesting in the light of the proposal that the dominant suppressor phenotype of *prp16-1* might be explained by a change in a proofreading mechanism²². Mutations that increase the rate of formation of a committed (productive) complex in a branched kinetic proofreading pathway should act to decrease fidelity by reducing the time available for dissociation of the incorrect complex via the 'discard' pathway. Such a model has been experimentally verified for the decoding of the ternary EF-Tu/tRNA/GTP complex at the ribosomal A site³⁷. We are now in a position to test whether the decreased fidelity of the *prp16-1* mutant is reflected in an altered affinity for, or rate of hydrolysis of, ATP in our *in vitro* assays.

PRP16 (ref. 22) was the first member of the putative RNA-dependent ATPase family²⁴ to be implicated in nuclear pre-mRNA splicing. Since this discovery the number of new family members has increased dramatically³⁸. In particular, PRP5 (ref. 39), PRP28 (E. Strauss and C.G., submitted) and SPP81 (D. Jamieson and J. Beggs, personal communication) contain good matches to each of the seven motifs proposed to define the

eIF-4A-like family and have been hypothesized to be ATP-dependent RNA helicases involved in nuclear splicing. Interestingly, PRP16, which was noted to contain variants of the eIF-4A-like motifs²², now appears to be the prototype of a related but distinct family. Recently, two other yeast splicing factors have been identified that share extensive homologies with PRP16 in a 320-amino-acid region encompassing the putative ATP-binding domain. PRP2 and PRP22 are ~50% identical to PRP16 in this region (amino acids 374-692). Notably, this identity is strongest for the sequences which distinguish the eIF-4A-like motifs from those of PRP16. Thus, for example, the sequence DEAD, the first aspartic acid residue of which is thought to interact with the phosphates of ATP via Mg²⁺ (ref. 43), is replaced in the PRP16 family by DEAH; the PTRELA motif 67 amino acids upstream is replaced by PRRVAA; and the sequence HRIGRTGR, found about 212 amino acids downstream of DEAD, is replaced by QRXGRAGR. The PTRELA and HRIGR motifs have been proposed to be involved in helicase function²⁴. Further experiments are required to determine the roles of these motifs in the DEAD and DEAH protein families and their relationship to one another.

In addition to the conserved 320-amino-acid central region, the C-terminal region of ~280 amino acids (693-975 in PRP16) is 42% and 35% identical to PRP22 and PRP2, respectively. The three DEAH-type proteins contain distinctive N-terminal sequences. It is striking that each of these family members appears to have a unique role in the splicing pathway. PRP2 is dispensable for 40S spliceosome assembly, but is required, together with ATP and at least one other factor, for the first cleavage-ligation reaction²⁰. We have shown that PRP16 is required for the second cleavage-ligation reaction. In contrast, PRP22 appears to promote a very late step, release of the spliced message from the spliceosome⁵¹. It seems likely that all three proteins are RNA-dependent ATPases, differing in their ligands and effector functions. It is of great interest to know whether PRP2 and PRP22, like PRP16, undergo transient high-affinity interactions with the spliceosome which are altered, either directly or indirectly, by the binding and/or hydrolysis of ATP. It is tempting to speculate that the PRP16-type proteins carry out a set of mechanistically related processes which are conceptually analogous to the function of ribosomal factors in protein synthesis^{41,42}, using nucleotide hydrolysis to drive a series of unidirectional conformational switches. □

Received 19 November; accepted 20 December 1990.

- Cech, T. R. & Bass, B. L. *A. Rev. Biochem.* **55**, 599-629 (1986).
- Sharp, P. A. *Science* **235**, 766-771 (1987).
- Guthrie, C. *Am. Zool.* **29**, 557-567 (1989).
- Michel, F., Umesono, K. & Ozeki, H. *Gene* **82**, 5-30 (1989).
- Green, M. R. *A. Rev. Genet.* **20**, 671-708 (1986).
- Woolford, J. L. *Jr Yeast* **5**, 435-457 (1989).
- Guthrie, C. & Patterson, B. *A. Rev. Genet.* **23**, 387-419 (1988).
- Lührmann, R. in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles* (ed. Birnstiel, M. L.), 71-99 (Springer, Berlin, 1988).
- Cheng, S. C. & Abelson, J. *Genes Dev.* **1**, 1014-1027 (1987).
- Konarska, M. M. & Sharp, P. A. *Cell* **49**, 763-774 (1987).
- Hartwell, L. H. *J. Bact.* **93**, 1662-1670 (1967).
- Vijayraghavan, U., Company, M. & Abelson, J. *Genes Dev.* **3**, 1206-1216 (1989).
- Lustig, A. J., Lin, R. J. & Abelson, J. *Cell* **47**, 953-963 (1986).
- Vijayraghavan, U. & Abelson, J. *Molec. cell. Biol.* **10**, 324-332 (1990).
- Lossky, M., Anderson, G. J., Jackson, S. P. & Beggs, J. *Cell* **51**, 1019-1026 (1987).
- Banroques, J. & Abelson, J. *Molec. cell. Biol.* **9**, 3710-3719 (1989).
- Björn, S. P., Soltysik, A., Beggs, J. D. & Friesen, J. D. *Mol. cell. Biol.* **9**, 3698-3709 (1989).
- Chang, T. H., Clark, M. W., Lustig, A. J., Cusick, M. E. & Abelson, J. *Molec. cell. Biol.* **8**, 2379-2393 (1988).
- Whittaker, E., Lossky, M. & Beggs, J. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2216-2219 (1990).
- Lin, R. J., Lustig, A. J. & Abelson, J. *Genes Dev.* **1**, 7-18 (1987).
- Couto, J. R., Tamm, J., Parker, R. & Guthrie, C. *Genes Dev.* **1**, 445-455 (1987).
- Burgess, S., Couto, J. R. & Guthrie, C. *Cell* **60**, 705-717 (1990).
- Rozen, F. *et al. Molec. cell. Biol.* **10**, 1134-1144 (1990).
- Linder, P. *et al. Nature* **337**, 121-122 (1989).
- Schena, M., Picard, D. & Yamamoto, K. R. *Meth. Enzym.* **194**, 389-398 (1990).
- Lin, R. J., Newman, A. J., Cheng, S. C. & Abelson, J. *J. Biol. Chem.* **260**, 14780-14792 (1985).
- England, T. E. & Uhlenbeck, O. C. *Nature* **275**, 560-561 (1978).

- Cheng, S. C. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2387-2391 (1986).
- MacArthur, H. & Walter, G. *J. Virol.* **52**, 483-491 (1984).
- Clark, R., Lane, D. P. & Tjian, R. *J. Biol. Chem.* **256**, 11854-11858 (1981).
- Grifo, J. A., Abramson, R. D., Sattler, C. A. & Merrick, W. C. *J. Biol. Chem.* **259**, 8648-8654 (1984).
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T. & Schrier, J. *Nature* **336**, 496-498 (1988).
- Hirling, H., Scheffner, M., Restle, T. & Stahl, H. *Nature* **339**, 562-564 (1989).
- Iggo, R. D. & Lane, D. P. *EMBO J.* **8**, 1827-1831 (1989).
- Sawa, H., Ohno, M., Sakamoto, H. & Shimura, Y. *Nucleic Acids Res.* **16**, 3157-3164 (1988).
- Sergeant, A., Bohnmann, D., Zentgraf, H., Weiher, H. & Keller, W. *J. molec. Biol.* **180**, 577-600 (1984).
- Thompson, R. C. *Trends biochem. Sci.* **13**, 91-93 (1988).
- Chang, T. H., Arenas, J. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1571-1575 (1990).
- Dalbadie-Mcfarland, G. & Abelson, J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4236-4240 (1990).
- Hodgman, T. C. *Nature* **333**, 22-23 (1988).
- Rhoads, R. E. *Trends biochem. Sci.* **13**, 52-56 (1988).
- Gualerzi, C. O. & Pon, C. L. *Biochemistry* **29**, 5881-5889 (1990).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
- Spindler, K. R., Rosser, D. S. E. & Berk, A. J. *J. Virol.* **49**, 132-141 (1984).
- Snyder, M. *J. Cell Biol.* **108**, 1419-1429 (1989).
- Ito, H., Fukuda, Y., Murata, K. & Kimura, A. *J. Bact.* **153**, 163-168 (1983).
- Wray, W., Boulikas, T., Wray, V. P. & Hancock, R. *Analyt. Biochem.* **118**, 197-203 (1981).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).
- Barnes, D. A. & Thorner, J. *Molec. cell. Biol.* **6**, 2828-2838 (1986).
- Hamm, J., Kazmaier, M. & Mattaj, J. W. *EMBO J.* **6**, 3479-3485 (1987).
- Company, M., Arenas, J. & Abelson, J. *Nature* **349**, 487-493 (1991).

ACKNOWLEDGEMENTS. We thank John Abelson, Mashid Company, Jaime Arenas, Ren-Jan Lin, Jiann-Hwa Chen, Jean Beggs and Derek Jamieson for communicating sequences and results prior to publication. John Abelson for the *prp23* strain, Lucita Esperas for technical assistance, and Peter Walter, Evelyn Strauss and other members of the laboratory for comments on the manuscript. This work was supported by the NIH. B. S. is an EMBO postdoctoral fellow.

Positive water vapour feedback in climate models confirmed by satellite data

D. Rind*, E.-W. Chiou†, W. Chu†, J. Larsen†, S. Oltmans‡, J. Lerner*, M. P. McCormick† & L. McMaster†

* NASA/Goddard Institute for Space Studies, New York, New York 10025, USA

† NASA/Langley Research Center, Hampton, Virginia 23665-5225, USA

‡ NOAA CMDL, Boulder, Colorado, 80302, USA

CHIEF among the mechanisms thought to amplify the global climate response to increased concentrations of trace gases is the atmospheric water vapour feedback. As the oceans and atmosphere warm, there is increased evaporation, and it has been generally thought that the additional moisture then adds to the greenhouse effect by trapping more infrared radiation. Recently, it has been suggested that general circulation models used for evaluating climate change overestimate this response, and that increased convection in a warmer climate would actually dry the middle and upper troposphere by means of associated compensatory subsidence¹. We use some new satellite-generated water vapour data to investigate this question. From a comparison of summer and winter moisture values in regions of the middle and upper troposphere that have previously been difficult to observe with confidence, we find that, as the hemispheres warm, increased convection leads to increased water vapour above 500 mbar in approximate quantitative agreement with the results from current climate models. The same conclusion is reached by comparing the tropical western and eastern Pacific regions. Thus, we conclude that the water vapour feedback is not overestimated in models and should amplify the climate response to increased trace-gas concentrations.

Studies with the Goddard Institute for Space Studies (GISS) general circulation model (GCM) show that, of the 4.2 °C warming that results from doubling atmospheric CO₂, ~1.7 °C is contributed by the increase in water vapour². In this regard, Lindzen¹ believes that the increased convection associated with the projected warming would act to dry the middle and upper troposphere. In his hypothesis, air rising in cumulus convective towers would cool and saturate, producing rain, which would

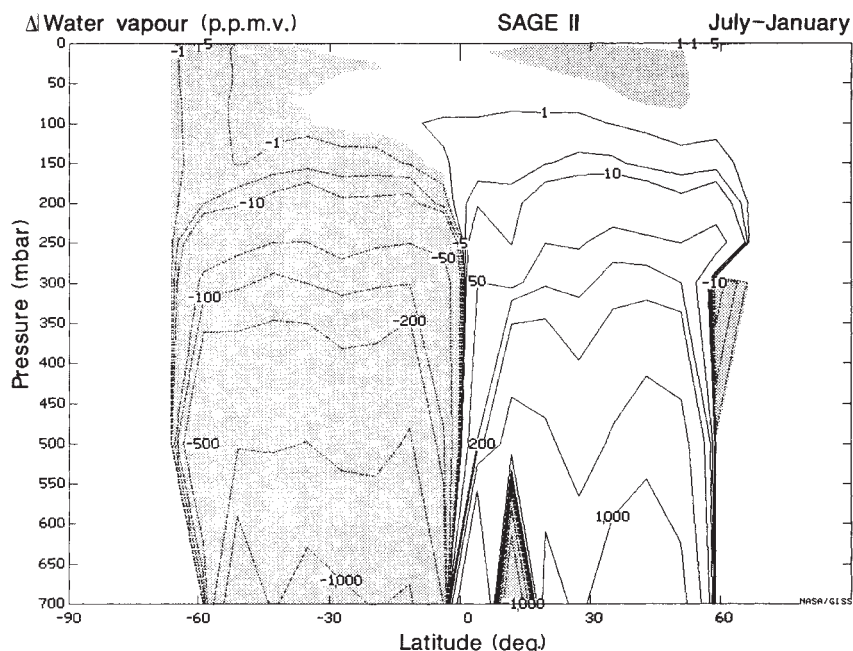
fall out of the column. To balance the upward air-mass flux, there would have to be compensating downward motion, or subsidence, which would dry the atmosphere above ~5 km, thereby leading to a negative water vapour feedback.

In the vicinity of warmer sea surface temperatures, there will be greater convection; indeed, satellite observations have been used to verify that the greenhouse effect is larger over warmer ocean regions³, suggesting that a positive feedback arises from increased convection. Here we address the question more directly by using satellite-derived water vapour observations to test the water vapour response to increased convection in regions of the middle and upper troposphere that have previously been difficult to observe with other methods.

The water vapour data set has been generated by the SAGE II (Stratospheric Aerosol and Gas Experiment) instrument aboard the Earth Radiation Budget Satellite⁴ (ERBS). SAGE II has been collecting data continuously since its launch in October 1984. It is a solar occultation instrument measuring attenuated sunlight through the Earth's limb with a path length of ~200 km and a vertical resolution of 1 km. Each sunrise and sunset experienced by the spacecraft results in a solar occultation measurement yielding ~900 opportunities to retrieve gas-concentration profiles. The path length is small enough to isolate individual areas, but large enough to encompass most of the observed vertical motions associated with convection¹⁰. Water vapour is retrieved by determining the extinction of the solar signal at 0.94 µm. The instrument is self-calibrating as it observes the unattenuated sun above the atmosphere before or after each event. The water vapour observations have been validated by comparison with radiosonde data, frost point hygrometer, Lyman-α and LIMS satellite observations⁵, and have been shown to produce realistic values, with an estimated accuracy of ~10%.

We compared the relative humidity calculated using the SAGE II water vapour data with that calculated using radiosonde observations with the Vaisala water vapour sensor, an instrument now in widespread use. The temperature data needed for the SAGE relative humidity calculations were provided by the National Weather Service of the National Oceanographic and Atmospheric Administration. The estimated radiosonde sensor accuracy is 2–3%, but this value becomes more uncertain with increasing altitude⁶. A comparison was made whenever the satellite and *in situ* measurements were within 250 km and 6 h during 1987.

FIG. 1 Change in water vapour between July and January from SAGE II observations for January 1985 to July 1989.



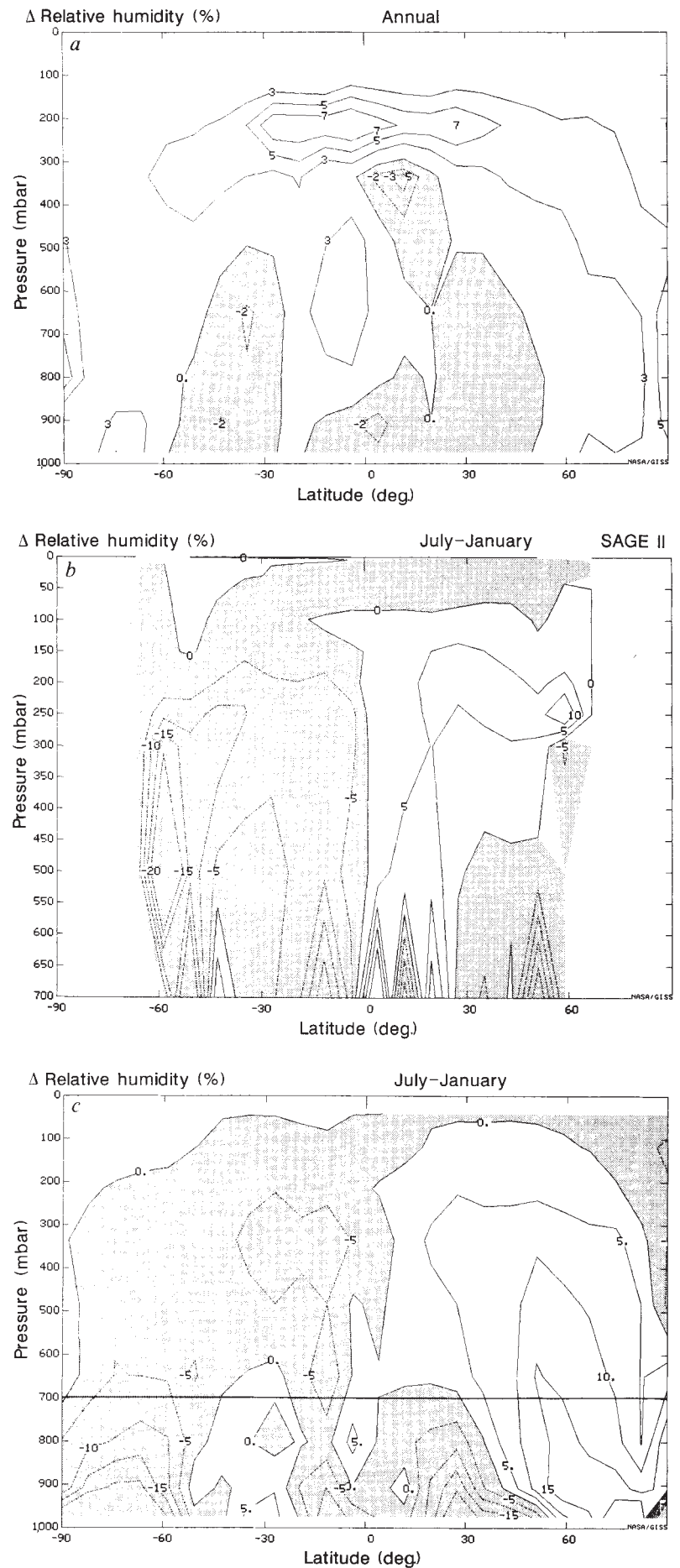


FIG. 2 Change of relative humidity *a*, between the doubled- CO_2 climate and the current climate in the GISS model; *b*, between July and January, SAGE II observations, January 1985 to July 1989; *c*, between July and January in the GISS GCM, for which three July and three January months were used from different years of a five-year simulation of the present climate. For comparison between model and observations, the levels above 700 mbar (solid line) should be considered.

TABLE 1 Comparison of SAGE II and radiosonde relative humidity

Altitude (km)	Number of comparisons	Average difference (%) SAGE II – sonde	Absolute difference (%) SAGE II – sonde
5.5	76	-4.3	14.1
6.5	119	-2.4	16.2
7.5	143	3.1	16.1
8.5	174	5.8	14.8
9.5	178	6.7	14.0
10.5	201	4.3	8.6
11.5	216	1.7	5.5
12.5	182	0.5	3.1

The results are shown in Table 1. The average difference between the two data sets is within the instrument uncertainties, and the SAGE data show no obvious bias. The slightly higher SAGE relative humidities in the upper troposphere may be the result of a systematic difference in the calculation of relative humidity at sub-freezing temperatures—with respect to ice in the satellite data set, and with respect to water in the radiosonde retrievals. For example, using global average values, this difference would artificially increase SAGE relative humidities by ~9% at 9.5 km. The absolute differences between the techniques are somewhat larger, possibly due to spatial and temporal heterogeneities that could not be removed.

To test the effects of convection on the atmospheric water vapour profile, we first examine the seasonal variation in troposphere moisture. As each hemisphere in summer experiences greater convection than in winter, we compared the SAGE II retrievals for each July and January for five years of data (Fig. 1). The specific humidity (mass of water vapour per unit mass of moist air) is consistently higher in July in the Northern Hemisphere, and January in the Southern Hemisphere, and there is clearly more water vapour in the middle and upper troposphere (~5–15 km) when there is increased convection. Latitude-longitude plots of specific humidity using the SAGE II data confirm that the specific humidity values are largest in the convective regions⁵.

It is also important to investigate the variation of specific humidity with temperature. Compared with the current climate, the GISS model calculates that in a climate with doubled CO₂, the relative humidity will remain approximately constant, with slight increases in the upper troposphere (Fig. 2a). Given that the model also predicts considerable temperature increases, of up to 8 °C in the upper troposphere, constant relative humidity implies a large increase in specific humidity, averaging ~33% with a 4 °C global warming.

To determine if the model is producing the proper quantitative water vapour change as a function of temperature under warming conditions, we compared the difference in relative humidity between summer and winter in the SAGE II observations to the results from the model simulation of the present climate (Fig. 2b, c). The SAGE II retrievals cannot be made through cloud cover, and so to allow the model to duplicate satellite-observing conditions, we reran three Januarys and three Julys from different years of the model simulations of the present climate. We then saved only the data that would have been observed by satellite (that is, we removed data below clouds). In both model and observations, data were available down to 5 km for ~50% of the time.

Both the SAGE II data and the model show that relative humidity does not change much between summer and winter in either hemisphere. There is a slight increase of ~5–10% in the convective (summer) hemisphere above 500 mbar in both the satellite data and the model, which is broadly similar to that forecast for the warmer climate (Fig. 2a). Although the differences between summer and winter are not precisely equivalent to the difference between the doubled-CO₂ climate and that of

today, both summer and the doubled-CO₂ climate feature warming and increased convection. For example, in the GISS GCM, the convective mass flux passing through 300 mbar in the Northern Hemisphere increases by 28% from winter to summer, whereas, by comparison, the global annual average convective mass flux through the same level increases by 10% in the doubled-CO₂ climate. The seasonal climate change is thus a more severe test for the effects of increased convection, which apparently does not result in drying of the middle and upper troposphere in either the SAGE II data or the model.

An additional test of the effects of convection can be made by investigating the humidity profiles in individual regions. Figure 3 shows the difference in relative humidity, calculated from the SAGE II data, between areas in the convective western Pacific and the largely non-convective eastern Pacific for the altitudes in question, between 5.5 and 11.5 km (~500–200 mbar). Results are shown for the two solstice seasons, when the two regions differ in precipitation by nearly 80% (ref. 7), almost all of which is convective in origin. The west Pacific has higher relative humidity throughout most of the altitude range, although the differences are small. At 400 mbar (~7.5 km), the observed temperatures in the western Pacific are nearly 2 °C warmer⁷, so that even if relative humidities were the same, the western Pacific would have ~20% greater specific humidity. Again, there is no indication of drying associated with increased convection. Shown for comparison are the model results, which also indicate a small increase in relative humidity. The model-generated change occurs with a somewhat different vertical profile. Although a detailed extrapolation is probably not warranted, it would appear that the model is underestimating the greenhouse effect of increased convection, as the predicted water vapour increase peaks at a lower altitude than in the SAGE II observations.

The observed moistening in both the seasonal and regional comparisons is apparently associated with increased convection, as it is most evident in the known convective regions of the western Pacific, and the Amazon and African rainforests during their rainy seasons⁵. Although increased subsidence would probably produce a drying tendency, other processes are affecting the moisture distribution, such as detrainment of water vapour at cloud top, and re-evaporation of moisture blown out of the convective plume. Furthermore, the large-scale circulation can respond to the convective heating with its own vertical transport mechanisms. In total, these other processes seem to offset the effects of drying due to subsidence.

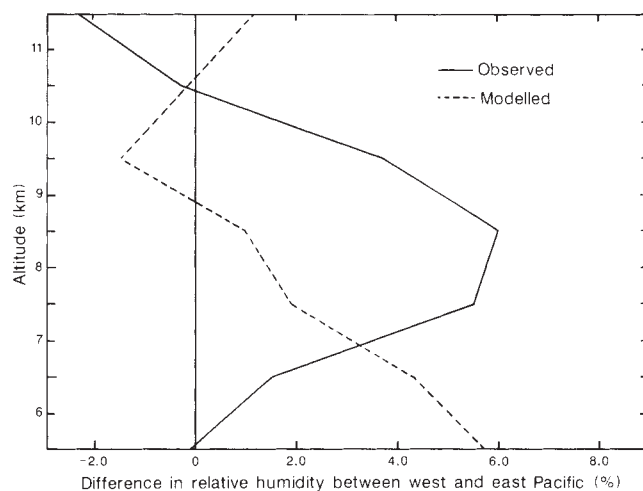


FIG. 3 Difference in relative humidity between the east Pacific (90°–120° W, 0°–30° N) and west Pacific (120°–150° E, 0°–30° N) for the two solstice seasons from the SAGE II data (observed) and the GISS GCM (modelled).

As noted by Cess⁸, the various GCMs agree remarkably well concerning clear-sky long-wave feedback, and the results of our study and that of Raval and Ramanathan³ indicate that GCMs appear to be consistent with two independent clear-sky data sets. The similarity of modelled and observed relative humidity variations under warming conditions is a first-order validation of the model's water vapour response, and does suggest that the models are not overestimating the water vapour feedback. The predicted 1.7 °C warming from this process, in combination with the 1.2 °C warming from doubled CO₂, brings the system response close to 3 °C. The ice-albedo feedback, due to decreased snow cover and sea ice, is also likely to be positive, with a magnitude, in the GISS GCM, of nearly 0.4 °C². That leaves only cloud properties as an obvious possible mitigating (or amplifying) factor. Given the potential impact of a 3–4 °C warming on droughts and ecosystems⁹, our results re-emphasize the importance of reducing trace-gas emissions. □

Received 26 November 1990; accepted 3 January 1991.

1. Lindzen, R. S. *Bull. Am. met. Soc.* **71**, 288–299 (1990).
2. Hansen, J. et al. *Climate Processes and Climate Sensitivity* (eds Hansen, J. & Takahashi, T.) 130–163 (Am. Geophys. Un., Washington, DC, 1984).
3. Raval, A. & Ramanathan, V. *Nature* **342**, 758–761 (1989).
4. Mauldin, L. E. III, Zaun, N. H., McCormick, M. P., Guy, J. H. & Vaughn, W. R. *Opt. Engng* **24**, 307–312 (1985).
5. Rind, D. et al. *J. geophys. Res.* (submitted).
6. Reeves, R., Williams, S., Rasmussen, E., Acheson, D., Carpenter, T. & Rasmussen, J. *NOAA tech. rep., EDS 20* (US Dept of Commerce, Washington, D.C., 1976).
7. Schutz, C. & Gates, W. L. *Rand Rep. R-915-ARPA, R-1029-ARPA*, Santa Monica, California, 1971, 1972.
8. Cess, R. D. *Nature* **342**, 736–737 (1989).
9. Rind, D., Goldberg, R., Hansen, J., Rosenzweig, C. & Ruedy, R. *J. geophys. Res.* **95**, 9983–10004 (1990).
10. Houze, R. A. Jr & Bates, A. K. *Rev. Geophys. Space Phys.* **19**, 541–576 (1981).

ACKNOWLEDGEMENTS. The SAGE II program and climate modelling at GISS are supported by the NASA Climate Program Office, whereas modelling the climate response to doubled CO₂ at GISS is supported by the United States Environmental Protection Agency.

Could reducing fossil-fuel emissions cause global warming?

T. M. L. Wigley

Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK

WHEN fossil fuel is burned, both carbon dioxide and sulphur dioxide are added to the atmosphere. The former should cause warming of the lower atmosphere by enhancing the greenhouse effect, whereas the latter, by producing sulphate aerosols, may cause a cooling effect. The possibility that these two processes could offset each other was suggested many years ago (see, for example, ref. 1), but during most of the intervening period, attention has focused on the greenhouse effect. Interest in tropospheric aerosols has, however, recently been rekindled by the realization that they may influence climate, not only through clear-sky radiative effects^{2–5}, but also by modifying cloud albedo^{6–8}. Here I examine the sensitivity of the climate system to simultaneous changes in SO₂ and CO₂ emissions, as might occur if controls were imposed on fossil-fuel use. Over the next 10–30 years, it is conceivable that the increased radiative forcing due to SO₂ concentration changes could more than offset reductions in radiative forcing due to reduced CO₂ emissions.

The relative emissions of SO₂ and CO₂ from fossil-fuel combustion depend on fuel type and emissions-control technologies, and so could change in the future. As a baseline case, however, I assume that emissions of both gases will parallel total fuel use, and consider total fossil-fuel use to change with a constant growth rate. Future SO₂ and CO₂ emissions are therefore both

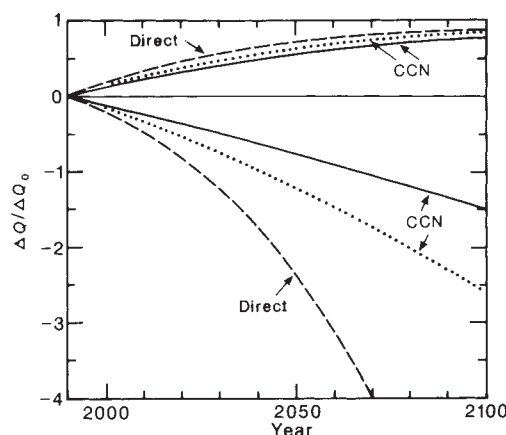


FIG. 1 Northern Hemisphere SO₂-related radiative-forcing changes (W m⁻²) for emissions growth rates of +2% yr⁻¹ (lower curves) and -2% yr⁻¹ (upper curves). The values are expressed relative to the appropriate forcing increment to 1990 ($\Delta Q_{0,dir}$ or $\Delta Q_{0,indir}$) with the sign changed so that the results reflect the sign of future changes (that is, a positive forcing effect for -2% yr⁻¹ growth rate). The uppermost and lowermost curves (long dashes) show the direct effect of changes in clear-sky sulphate aerosol. The innermost curves (full lines) show the indirect (CCN) effect for a present to pre-industrial SO₂ emissions ratio (E_1/E_0) of 3.6, as suggested by the modelling study of Charlson et al. (ref. 5). The short dashed curves show the indirect (CCN) effect for $E_1/E_0 = 1.8$. Note that the apparently larger CCN effect for smaller E_1/E_0 is only relative to $\Delta Q_{0,indir}$.

determined by an equation of the form

$$E = E_0 + (E_1 - E_0) e^{\alpha t} \quad (1)$$

where E_0 is the natural background emission rate, E_1 is the current total emission rate (at time $t = 0$) and α is the growth rate. To determine the possible effect of reductions in fossil-fuel use on climate, I will compare the SO₂- and CO₂-derived changes in radiative forcing for three scenarios: growth rates of zero and $\pm 2\% \text{ yr}^{-1}$.

Consider the direct SO₂ effect first. The forcing change, dQ_{dir} , is approximately proportional to the change in the integrated column burden of sulphate aerosol, which in turn is approximately proportional to the change in mean-column SO₂ concentration⁵. As the lifetime for SO₂ is much less than a year, changes in SO₂ concentration must parallel emission changes. The direct radiative forcing may therefore be considered as linearly related to changes in emissions:

$$dQ_{dir} = a dE \quad (2)$$

For the indirect SO₂ effect through changes in cloud albedo, $dQ_{indir} \propto dC/C$, where C denotes the cloud-condensation-nuclei (CCN) concentration^{6,7}. If changes in C are linearly related to changes in aerosol concentration (as assumed, for example, in refs 8 and 9) then, as changes in aerosol concentration should be approximately linearly related to changes in SO₂ emissions (see above), we obtain

$$dQ_{indir} \propto dE/E = b dE/E \quad (3)$$

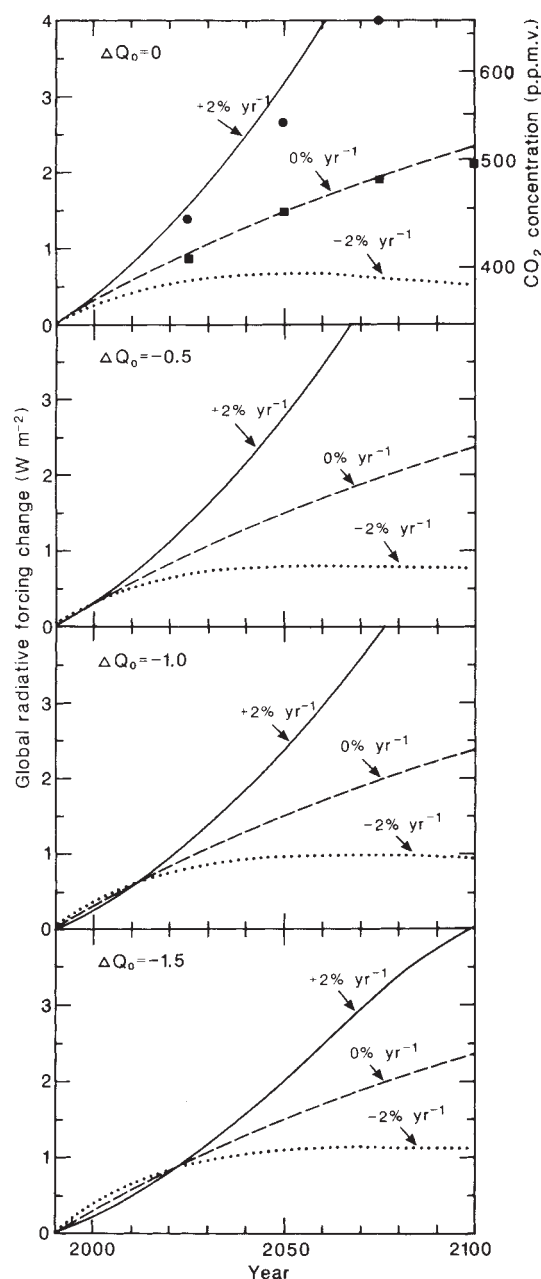
for the indirect forcing change. The C - A link (A denotes aerosol concentration) is known to be nonlinear^{10,11}, but equation (3) still applies provided that the relationship can be expressed in the form $C \propto A^n$.

The above relationships between emissions and forcing are only approximate, and the constants a and, in particular, b are of uncertain magnitude. These uncertainties can be circumvented by expressing future forcing changes in terms of the present man-made SO₂-related forcing (denoted by ΔQ_0). ΔQ_0 is not known, but the range of possible values can be estimated by noting that the preponderance of SO₂ emissions in the Northern Hemisphere should have led to a difference in the temperature record of the two hemispheres. As no striking difference is

TABLE 1 Combined CO₂ and aerosol forcing changes in the Northern Hemisphere for different emissions growth rates and different strengths of the aerosol effect (ΔQ_0)

Emissions growth rate (α)	ΔQ_0 (W m ⁻²)	2000	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
+2% yr ⁻¹	0.0	0.36	0.77	1.25	1.80	2.43	3.14	3.94	4.84	5.74	6.69	7.64
	-0.5	0.27	0.59	0.95	1.37	1.84	2.36	2.95	3.59	4.22	4.75	5.26
	-1.0	0.19	0.41	0.66	0.94	1.25	1.59	1.96	2.34	2.66	2.82	2.88
	-1.5	0.10	0.22	0.36	0.51	0.66	0.82	0.97	1.09	1.10	0.89	0.50
0% yr ⁻¹	all	0.30	0.58	0.83	1.06	1.27	1.47	1.67	1.85	2.03	2.20	2.37
-2% yr ⁻¹	0.0	0.26	0.43	0.53	0.59	0.63	0.64	0.63	0.61	0.58	0.55	0.51
	-0.5	0.33	0.56	0.72	0.83	0.90	0.95	0.97	0.98	0.97	0.96	0.94
	-1.0	0.40	0.70	0.91	1.07	1.18	1.26	1.31	1.35	1.36	1.37	1.36
	-1.5	0.48	0.83	1.10	1.30	1.46	1.57	1.65	1.71	1.75	1.77	1.78

For $\alpha = 0\% \text{ yr}^{-1}$ (middle panel) the aerosol effect is zero and the values shown are therefore those for CO₂ alone. Similarly, for any α value, the aerosol effect is zero for $\Delta Q_0 = 0 \text{ W m}^{-2}$. These rows therefore give the effect of CO₂ alone for $\alpha \neq 0\% \text{ yr}^{-1}$. They also give the changes assumed for the Southern Hemisphere.



evident, this puts an upper limit on the magnitude of ΔQ_0 (ref. 8). Figure 1 of ref. 8 shows that, for the Northern Hemisphere, $0 \geq \Delta Q_0 \geq -1.5 \text{ W m}^{-2}$ is the most likely range. For the Southern Hemisphere, δQ_0 is much smaller and is assumed to be zero. If ΔQ is the forcing change from today, then equations (1) and (2) yield

$$\delta Q_{\text{dir}} = \Delta Q_{0,\text{dir}}(e^{a t} - 1) \quad (4)$$

and equations (2) and (3) give

$$\delta Q_{\text{indir}} = \Delta Q_{0,\text{indir}} \frac{\ln [E_0/E_1 + (1 - E_0/E_1) e^{a t}]}{\ln [E_1/E_0]} \quad (5)$$

The unknowns a and b have been absorbed into ΔQ_0 .

These results show that, if α is negative (that is, if there were a reduction in emissions, eventually back to the natural background level E_0), then δQ will tend asymptotically towards $-\Delta Q_0$ (as it obviously must). The relative behaviour of the direct and indirect effects is shown in Fig. 1. For declining emissions (upper part of Fig. 1), the two effects are comparable (with a slightly longer timescale for the indirect effect). For increasing emissions, because of the nonlinear link between emissions and CCN-related radiative forcing, the direct effect dominates (relative to the appropriate 1990 level).

For the radiative-forcing effect of CO₂ emission changes, changes in CO₂ concentration for the emissions scenarios described by equation (1) must be estimated. I have done this using the convolution form of the carbon cycle model of Maier-Reimer and Hasselmann^{12,13}. To reproduce past concentration changes using this model, the CO₂ emissions must rise from a negligible pre-industrial value to 6.18 Gt C in 1990. Future emissions were therefore assumed to be of the form

$$E = 6.18 e^{a t}$$

FIG. 2 Global radiative forcing changes (in W m⁻² from 1990) due to the combined effects of CO₂ and SO₂ emissions for growth rates of +2% yr⁻¹ (full lines in all panels), 0% yr⁻¹ (long dashed lines) and -2% yr⁻¹ (short dashed lines). The upper panel gives the results for CO₂ alone ($\Delta Q_0 = 0$). Here, the right-hand scale shows CO₂ concentration (in ppmv). For comparison, concentration changes for the IPCC 'Business as Usual' scenario and IPCC scenario C are shown (dots and squares respectively). The former corresponds to a situation in which future energy supply is coal-intensive and only modest efficiency increases are achieved, whereas the latter represents the imposition of strong controls on fossil-fuel emissions, through large efficiency increases and a shift towards renewable and/or nuclear energy sources. The lower panels show the forcing changes for increasing magnitudes of the aerosol effect ($\Delta Q_0 = -0.5, -1.0, -1.5 \text{ W m}^{-2}$). The full dashed curve (0% yr⁻¹ growth rate) is the same in each panel because the aerosol effect is zero in this case, irrespective of ΔQ_0 . In all cases, there is a net increase in radiative forcing. If the aerosol effect is large (see especially the lower panels), the forcing increase is initially greatest for the lowest, negative growth rate.

consistent with equation (1). (Strictly, fossil fuel and deforestation sources should be considered separately, but this makes little difference to the results.) Computed concentration changes are shown in the top panel of Fig. 2 (right-hand scale).

For $\alpha < 0$, there is a striking contrast between the CO_2 behaviour and the SO_2 -related changes in forcing. Because of the long timescales involved in the carbon cycle, the response of CO_2 concentrations to a decline in emissions is very slow—in fact, for $\alpha = -2\%$, it is not until around 2050 that CO_2 concentrations begin to decline, by which time emissions have dropped to 30% of their 1990 level. SO_2 responses, however, are immediate. The initial response of the climate system to parallel reductions in the growth of CO_2 and SO_2 emissions would, therefore, be a reduced warming rate due to the slow-down in the rise in CO_2 concentration, offset by an increased warming rate due to

the effect of sulphate-aerosol changes¹⁴. Which of these two processes is likely to dominate?

To answer this question, I consider the combined radiative-forcing changes due to simultaneous changes in CO_2 and SO_2 emissions. CO_2 -forcing changes can be calculated using the above CO_2 -concentration scenarios and the forcing-concentration relationship used by Working Group 1 of the Intergovernmental Panel on Climate Change (IPCC)¹⁵, namely, $\delta Q = 6.3 \ln(C/C_0)$ for a concentration change from C_0 to C .

Because the relative importance of the direct and indirect effects of sulphate aerosols is not known, I have assumed them to be of equal importance ($\Delta Q_{0,\text{dir}} = \Delta Q_{0,\text{indir}}$) with ΔQ_0 , their sum, being between 0 and -1.5 W m^{-2} in the Northern Hemisphere and zero in the Southern Hemisphere. For negative growth rates of future SO_2 emissions, it makes little difference how the two effects are weighted, and, for positive growth rates, only minor quantitative differences follow for different weightings, at least until 2050. As the magnitudes of both sulphate-aerosol effects are uncertain, I have calculated future forcing changes for four different values of the Northern Hemisphere total forcing: 0.0, -0.5 , -1.0 and -1.5 W m^{-2} . The lowest value gives the base case of no-aerosol effect, whereas the largest value corresponds to a situation where $\sim 60\%$ of the Northern Hemisphere enhanced greenhouse forcing has been offset by SO_2 -related processes. The sulphate-aerosol effect is applied only in the Northern Hemisphere. The total (aerosol plus CO_2) Northern Hemisphere forcing changes are listed in Table 1, whereas the global forcing changes are shown in Fig. 2.

The primary effect of the aerosol contribution is to reduce the change in total forcing that would accrue from an emissions control policy. This narrows the range of forcing possibilities in the future. Over the period 1990–2050, for example, the range $\alpha = -2\%$ to $+2\%$ gives global-mean forcing increases in the range 0.64 W m^{-2} to 3.14 W m^{-2} in the absence of an aerosol effect, but a range of only 1.10 W m^{-2} to 1.98 W m^{-2} with a strong aerosol effect ($\Delta Q_0 = -1.5 \text{ W m}^{-2}$). Over the next 10–30 years, it is the aerosol effect that will chiefly influence changes in total forcing, to the extent that reduced emissions produce a larger forcing increase (and hence a larger global warming potential) than increased emissions.

Figure 2 shows the combined forcing for future emissions scenarios, but the effect of a policy to reduce emissions is better illustrated by considering the differences between these scenarios. The $\alpha = +2\%$ case characterizes the situation where future growth continues at the present rate. Modelled CO_2 concentrations in this case are very similar to those for the IPCC 'Business as Usual' scenario¹⁵ out to 2020 (Fig. 2, top panel). $\alpha = 0\%$ gives future CO_2 concentrations very similar to IPCC scenario C, and represents a control scenario suggested by a number of developed countries. $\alpha = -2\%$ represents a more extreme control scenario, at the limits of viability.

In Fig. 3, the effects of parallel reductions in the growth rates of CO_2 and SO_2 emissions are shown by plotting the forcing changes for the two 'emissions control' cases (namely, $\alpha = 0\%$ and $\alpha = -2\%$) relative to the no control, 'Business as Usual' case ($\alpha = +2\%$). Even with no aerosol effect, it is several decades before any noticeable radiative-forcing response to an emissions reduction policy occurs, simply because the carbon cycle has such a long response time. The aerosol effect delays the response still further, completely offsetting it for up to three decades. The emissions control cases actually show an increase in forcing (that is, a relative warming), the duration and strength of which depends on the strength of the aerosol effect (specified here by ΔQ_0). For a strong aerosol effect ($\Delta Q_0 = -1.5 \text{ W m}^{-2}$) there is a relative warming out to beyond 2020. This tendency would be exacerbated by any SO_2 -specific emissions control measures and/or if CO_2 emissions reductions were preferentially made in high-sulphur fuels, and reduced by CO_2 reductions from non-fossil-fuel sources.

The sulphate-aerosol effect is subject to large uncertainties^{4,16},

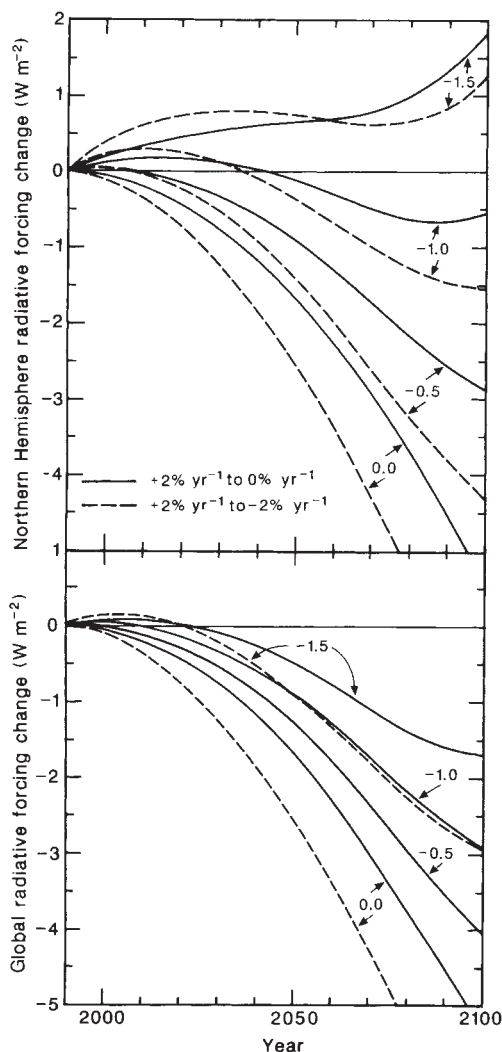


FIG. 3 Radiative forcing changes due to parallel reductions in CO_2 and SO_2 emissions growth rates from $+2\% \text{ yr}^{-1}$ to $0\% \text{ yr}^{-1}$ (full lines) or from $+2\% \text{ yr}^{-1}$ to $-2\% \text{ yr}^{-1}$ (dashed lines). The upper panel gives the results for the Northern Hemisphere, whereas the lower panel shows the global changes. The results are shown for four values of the parameter ΔQ_0 (W m^{-2}), which determines the strength of the SO_2 -derived aerosol effect. $\Delta Q_0 = 0$ (lower pairs of curves in both panels) gives the effects of reductions in CO_2 alone with no aerosol effect, which are the same in both hemispheres and hence in both panels. For $\Delta Q_0 = 0$, the effect of a reduced emissions growth rate is a cooling effect (a negative forcing change). As ΔQ_0 increases in magnitude, the cooling effect becomes smaller. For a large aerosol effect ($\Delta Q_0 = -1.5 \text{ W m}^{-2}$), the overall effect of the emissions reductions changes sign and yields a pronounced warming effect in the Northern Hemisphere, which translates to a global warming effect out to around 2020. (For simplicity, the large-emissions change case—dashed lines—is shown only for $\Delta Q_0 = 0$ and -1.5 W m^{-2} in the lower panel.)

accounted for here by taking a range of ΔQ_0 values. As shown in Fig. 3, these uncertainties are such that we cannot even predict the sign of the forcing changes that would result from controls on fossil-fuel use¹⁴. Even a small aerosol effect, however, would delay the response of the climate system to future attempts to control or limit the greenhouse problem.

It would, of course, be entirely wrong to consider SO_2 -related cooling in isolation as a benefit. In global-mean terms, CO_2 forcing could be markedly offset by the effect of sulphate-aerosol increases, but continued growth in emissions would lead to a rapidly widening contrast in interhemispheric forcing, potentially even more disruptive to the climate system than a uniformly distributed 'pure' greenhouse effect. Rather than acting as a panacea, the possible effects of fossil-fuel-derived sulphate aerosols should be seen as further reason for implementing controls on fossil-fuel use. □

Received 20 August 1990; accepted 3 January 1991.

1. Mitchell, J. M. Jr in *The Changing Global Environment* (ed. Singer, S. F.) 149–173 (Reidel, Dordrecht, 1975).
2. Rasool, S. I. & Schneider, S. H. *Science* **173**, 138–141 (1971).
3. Hansen, J. E., Lacis, A. A., Lee, P. & Wang, W.-C. *Ann. N.Y. Acad. Sci.* **338**, 575–587 (1980).
4. Charlson, R. J., Langner, J. & Rodhe, H. *Nature* **348**, 22 (1990).
5. Charlson, R. J., Langner, J. & Rodhe, H. *Tellus* (in the press).
6. Twomey, S. A., Peipgrass, M. & Wolfe, T. L. *Tellus* **36B**, 356–366 (1984).
7. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
8. Wigley, T. M. L. *Nature* **339**, 365–367 (1989).
9. Legrand, M. E., Delmas, R. J. & Charlson, R. J. *Nature* **334**, 418–420 (1988).
10. Leaitch, W. R., Strapp, J. W., Isaac, G. A. & Hudson, J. G. *Tellus* **38B**, 328–344 (1986).
11. Baker, M. B. & Charlson, R. J. *Nature* **345**, 142–145 (1990).
12. Maier-Reimer, E. & Hasselmann, K. *Clim. Dyn.* **2**, 63–90 (1987).
13. Harvey, L. D. D. *Climatic Change* **15**, 343–381 (1989).
14. Hansen, J. E. & Lacis, A. A. *Nature* **346**, 713–719 (1990).
15. Shine, K. P., Derwent, R. G., Wuebbles, D. J. & Morcrette, J.-J. in *Climate Change, The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 41–68 (Cambridge University Press, 1990).
16. Albrecht, B. A. *Science* **245**, 1227–1230 (1989).

ACKNOWLEDGEMENTS. Comments from Dan Lashof and Mike MacCracken helped to improve the manuscript. This work was supported by the United States Department of Energy, Atmospheric and Climate Research Division.

Geometric phase shifts in chemical oscillators

Michael L. Kagan*, Thomas B. Kepler†
& Irving R. Epstein‡

* Department of Chemistry, † Department of Biology and
‡ Centre for Complex Systems, Brandeis University,
Waltham, Massachusetts 02254-9110, USA

ONE of the most remarkable developments in quantum mechanics in recent years has been the discovery that when a system is moved adiabatically around a closed loop in parameter space there occurs, besides the familiar dynamical phase shift, an additional phase shift (sometimes referred to as 'Berry's phase') that is purely geometric in nature^{1–3}. The dynamical phase shift, which results from the variation of the period of the oscillatory system with the change in parameters, is relatively easily understood and is proportional to the time over which the parameter change occurs. The geometric phase shift, on the other hand, is less intuitive and depends on the curvature of the surface in parameter space bounded by the closed path, but is independent of the time taken to traverse the circuit. Here we present evidence for time-independent geometric phase shifts in numerical solutions for a model of an oscillating chemical reaction. The conditions for the occurrence of such shifts seem to be sufficiently general that geometric phase effects should be experimentally observable in essentially all chemical oscillators as well as in biological networks such as the brain and the central nervous system, where phase control is of vital importance⁴.

A phase shift analogous to Berry's phase—the Hannay

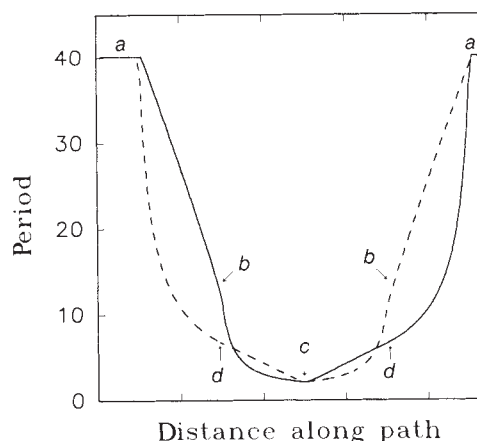


FIG. 1 Change in the period of the model oscillator (equations (3) and (4)) as the parameters are varied along a rectangular path defined by the vertices $(A, B) = (50, 0.5)(a)$, $(12, 0.5)(b)$, $(12, 5)(c)$, $(50, 5)(d)$. Heavy curve shows clockwise path; light curve shows anticlockwise path.

angle⁵—exists for classical mechanical hamiltonian (that is, conservative) systems, and the semi-classical connection between the two has been developed by Berry⁶. Examples of geometric phase shifts have been identified in quantum mechanics⁷, lasers^{8,9}, nuclear magnetic resonance¹⁰, gauge kinematics¹¹ and hydrodynamics¹². Recently Kepler and Kagan¹³ have studied phase shifts in non-hamiltonian (that is, dissipative) systems. They have shown that geometric phase shifts occur for classical dissipative systems that are transported adiabatically around a closed loop in parameter space. An example of such a system is an oscillating chemical reaction, in which the control parameters may be varied slowly so as to traverse a cyclic path.

Chemical oscillators, of which the Belousov-Zhabotinsky reaction is the most well documented¹⁴, involve highly nonlinear kinetics in the form of feedback reactions (autocatalysis or inhibition) that result in periodic changes in concentrations. In continuous-flow stirred-tank reactors (CSTRs), such reactions can sustain stable oscillations of constant period and amplitude. The time evolution of at least two of the oscillating species in a reaction system traces out closed loops or attractors in phase space for a particular set of control parameters. A change in a control parameter, such as the flow rate or the concentration of

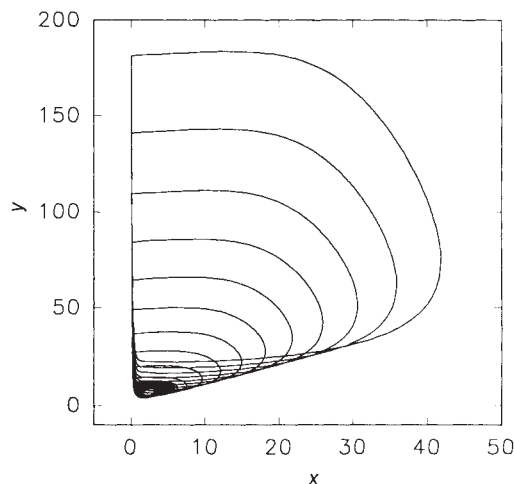


FIG. 2 Changing dimensions of the limit cycle when the parameters A and B are varied from $A=50$, $B=0.5$ (outermost cycle) to $A=12$, $B=5$ (innermost cycle) as in Fig. 1. The variables x and y are defined in equation (5).

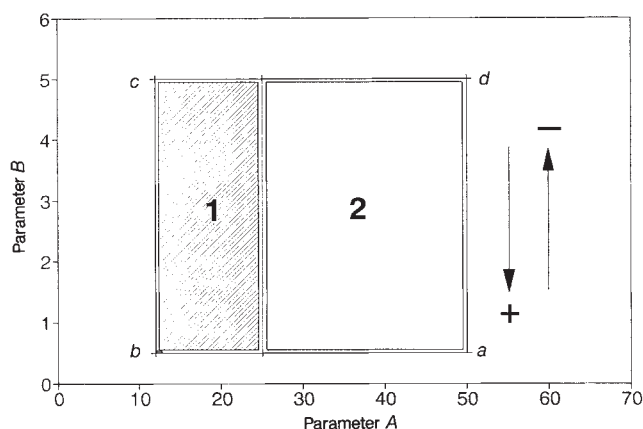


FIG. 3 Rectangular paths examined in the (A, B) parameter space. Path 3 is the full rectangular path used in Figs 1 and 2.

a species flowing into the reactor, generally results in changes in the period of the oscillations (Fig. 1) and in the position and shape of the attractor in phase space (Fig. 2).

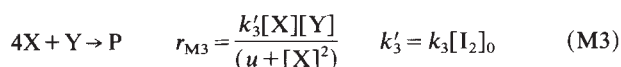
According to the formalism of ref. 13, if an oscillating reaction is taken adiabatically around a closed loop in parameter space, we should observe a phase shift composed of a dynamical, time-dependent component, a geometric, time-independent component and a non-adiabatic component that decays to zero as the time for completing the loop increases. If, starting from the same initial phase, the parameter circuit is traversed first in one direction and then in the other, the dynamical component will remain unchanged, whereas the geometric component will change sign (equation (4) of ref. 13). The non-adiabatic component can be eliminated only in the adiabatic limit, because it has no simple symmetries under time reversal. This consideration leads us to study the variable $\Delta\phi$, given by half the difference of the two overall phases:

$$\Delta\phi = (\varphi_{\text{tot}}^+ - \varphi_{\text{tot}}^-)/2 \quad (1)$$

$$\varphi_{\text{tot}}^{+(-)} = t_n^{+(-)}/\lambda_0 \quad (2)$$

where $\varphi_{\text{tot}}^{+(-)}$ is the total phase shift (dynamic + geometric + non-adiabatic) when the parameter space is traversed forwards (+) or in reverse (-), t_n is the time of the n th peak (one may choose any peak that occurs after the loop around the parameter space has been completed and the system has resumed oscillation at the initial period), and λ_0 is the period of the oscillator for the initial parameters. In the adiabatic limit, $\Delta\phi$ tends to the geometric phase shift.

Both temporal oscillations and spatial pattern formation in the chlorite-iodide-malonic acid (MA) oscillating reaction¹⁵ are described remarkably well by a simplified two-variable model¹⁶:



where the subscript 0 signifies concentrations that may be taken to be constant either as a result of the flow in a CSTR or because the oscillations occur during an early far-from-equilibrium stage of the batch reaction, and u is a parameter that describes the substrate inhibition in the chlorite-iodide reaction. Equations (M1-M3) yield the following system of dimensionless rate equations¹⁶:

$$dx/d\tau = A - x - 4xy/(1 + x^2) \quad (3)$$

$$dy/d\tau = B[x - xy/(1 + x^2)] \quad (4)$$

where

$$[X] = \alpha x, \quad [Y] = \beta y, \quad t = \gamma\tau, \quad \alpha = u^{1/2}, \quad (5)$$

$$\beta = uk'_2/k'_3 \quad \gamma = 1/k'_2, \quad A = k'_1\gamma/\alpha, \quad B = \alpha/\beta$$

To test for phase shifts, we varied the parameters A and B continuously around rectangular loops in parameter space as shown in Fig. 3. For a dissipative system described by an attractor, the analogue of the adiabatic theorem is that the time for relaxation back to the attractor after a small perturbation be negligible with respect to the timescale on which the parameters are varied. To ensure that initial transients had died out, the simulation was run for 500 time units before the parameters were varied. The simulations were carried out using a fourth-order Runge-Kutta integrator with adaptive step size.

Figure 4 shows $\Delta\phi$ for each of the paths in Fig. 3 as the total time for the parameter circuit is increased—that is, as we approach true adiabaticity. The fluctuations at short times result from non-adiabatic effects. At longer times the geometric phase shift tends to a constant value, in agreement with the simulations reported by Kepler and Kagan¹³ for an abstract model system. The dashed lines show the sum of the two sub-regions 1 and 2, indicating that the geometric phase shifts in the parameter space are additive. (We have not calculated $\Delta\phi$ using the theory of ref. 13 because such a calculation requires explicit parametrization of the limit cycle and yields a result that is convergent only asymptotically. The 'experimental' approach applied here is both more accurate and less costly.)

We have demonstrated that geometric phase shifts occur for a realistic model of an oscillating chemical reaction. Thus the notion of geometric phase shifts can be extended from Berry's phase as applied to quantum mechanics and Hannay's angle for integrable classical systems to classical dissipative systems. The model studied here is by no means unique; the nonlinearities found in any chemical oscillator should be sufficient to generate a geometric phase shift. Our preliminary calculations on the Brusselator¹⁷ model show an effect similar in magnitude to that found here.

The parameter variations carried out for our model equations could be realized experimentally by varying the input concentrations of two reactants, such as malonic acid and iodine, to a CSTR. The phase ϕ is easily monitored, for example as a fraction of the distance between peaks in the redox potential of the system (see equation (2)). The CSTR, a standard tool for the

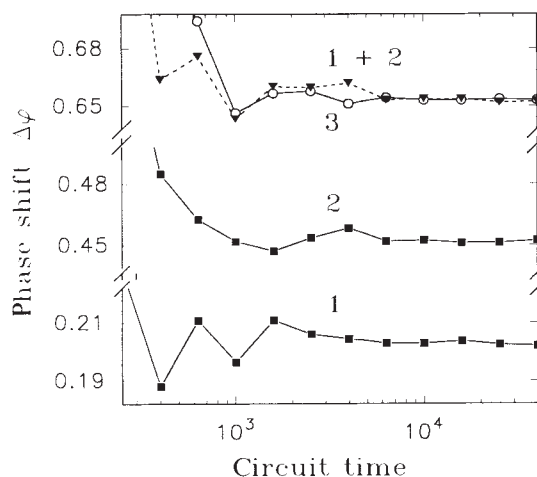


FIG. 4 Phase shifts $\Delta\phi$ for the parameter paths of Fig. 3: 1 is $A:25-12$, $B:0.5-5$; 2 is $A:50-25$, $B:0.5-5$; 3 is $A:50-12$, $B:0.5-5$; and the dashed line is the sum of the two smaller rectangular paths (1+2). Geometric phase shifts tend to the adiabatic limit on the right. The abscissa shows the total time taken to traverse path 3. For path 3, vertical segments are traversed in the same time as horizontal segments. For paths 1 and 2, parameters are varied at same rate as for path 3.

study of nonlinear chemical dynamics¹⁸, is ideally suited for experimental probes of geometric phase shifts in chemical oscillators, because one may easily vary such parameters as the flow rate, temperature and/or input concentrations in a cyclic fashion. We are now performing experiments of this type. □

Received 5 November; accepted 20 December 1990.

- Berry, M. V. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
- Shapere, A. & Wilczek, F. (eds) *Geometric Phases in Physics* (World Scientific, Singapore, 1989).
- Berry, M. V. *Sci. Am.* **259**(6), 46–52 (1988).
- Basar, E. & Bullock, T. (eds) *Induced Rhythms in the Brain* (Birkhauser, Boston, in the press).
- Hannay, J. H. *J. Phys. A* **18**, 221–230 (1985).
- Berry, M. V. *J. Phys. A* **18**, 15–27 (1985).
- Berry, M. V. *Proc. R. Soc. Lond. A* **414**, 31–46 (1987).
- Simon, R., Kimble, H. J. & Sudarshan, E. C. G. *Phys. Rev. Lett.* **61**, 19–22 (1988).
- Maddox, J. *Nature* **334**, 99 (1988).
- Suter, D., Pines, A. & Mueller, K. T. *Phys. Rev. Lett.* **60**, 1218–1220 (1988).
- Zee, A. *Phys. Rev. A* **38**, 1–6 (1988).
- Shapere, A. & Wilczek, F. *J. Fluid Mech.* **198**, 557–585 (1989).
- Kepler, T. B. & Kagan, M. L. *Phys. Rev. Lett.* (in the press).
- Field, R. J. & Burger, M. (eds) *Oscillations and Traveling Waves in Chemical Systems* (Wiley, New York, 1985).
- De Kepper, P., Epstein, I. R., Kustin, K., & Orbán, M. *J. phys. Chem.* **86**, 170–171 (1982).
- Lengyel, I., Rabai, G. & Epstein, I. R. *J. Am. chem. Soc.* **112**, 9104–9110 (1990).
- Prigogine, I. & Lefever, R. *J. chem. Phys.* **48**, 1695–1700 (1968).
- Epstein, I. R. *J. chem. Educ.* **66**, 191–195 (1989).

ACKNOWLEDGEMENTS. This work was supported by grants from the National Institute of Mental Health and the NSF. We thank Hugh Pendleton for helpful discussions and István Lengyel for computing assistance.

Low-temperature synthesis of hydrated zinco(beryllo)-phosphate and arsenate molecular sieves

Thurman E. Gier* & Galen D. Stucky†

* Central Research and Development Department, Experimental Station, E. I. duPont de Nemours, Inc., Wilmington, Delaware 19880-0262, USA

† Department of Chemistry, University of California at Santa Barbara, Santa Barbara, California 93106, USA

ALUMINOSILICATE (zeolite) molecular sieves, both natural and synthetic, have been studied extensively because of their utility in commercial processes such as petroleum cracking, water treatment and gas absorption^{1–3}. Although aluminium and silicon may be replaced by a variety of other tetrahedrally coordinated elements, there are few examples of molecular sieves that contain neither aluminium nor silicon: several gallophosphates are known^{4,5}, and cacoxenite, a basic ferric oxyphosphate, occurs naturally⁶. Recently, two new beryllophosphate minerals have been identified^{7,8}: tiptopite, isotypic with cancrinite, and pahasapaite (Li, Ca beryllophosphate), which has the structure of zeolite RHO. Harvey and Meier⁹ have prepared five new beryllophosphates, with structures analogous to RHO (Li), gismondine (Na), edingtonite (K), pollucite (Cs), and a new structure, BPH, that has no aluminosilicate analogue. Here we describe the preparation of new families of hydrated zincophosphates/arsenates and beryllophosphates/arsenates with structures that are also related to zeolitic aluminosilicates. These materials may be prepared from gels over a much wider range of pH and at much lower temperatures than are possible for aluminophosphates, perhaps because of the greater solubility of the framework elements: analogues of hydro-sodalite and zeolites RHO, Li-A(BW) and X are easily prepared at pH values ranging from 2 to 12 and between 4 and 100 °C. Growth of crystals adequate for X-ray structural analyses seems to be easier than is the case for aluminosilicate/phosphate materials.

Sodalites. We have prepared Na₆(H₂O)₈(ZnPO₄)₆ (DPZ-7A) and the corresponding arsenate (DPZ-7B), analogues of the hydro-sodalite¹⁰ Na₆(H₂O)₈(AlSiO₄)₆, at temperatures of 70 °C and below. Typical syntheses consist of dissolving 24 mmol

TABLE 1 Sodalites

<i>hkl</i>	Na ₆ (H ₂ O) ₈ (ZnPO ₄) ₆ (<i>a</i> = 8.824 Å)		Na ₆ (H ₂ O) ₈ (ZnAsO ₄) ₆ (<i>a</i> = 9.032 Å)	
	<i>D</i> _{obs}	Relative intensity	<i>D</i> _{obs}	Relative intensity
110	6.227	87	6.398	98
200	4.409	24	4.518	27
210	3.945	13	4.039	8
211	3.600	100	3.687	98
220	3.143	40	3.195	30
310	2.789	90	2.856	72
222	2.547	94	2.607	100
320	2.447	10	2.505	1
321	2.359	31	2.414	36
400	—	—	2.258	4
330	2.081	15	2.129	17
420	1.972	10	2.019	11
421	1.925	14	—	—
332	1.881	9	1.926	12
422	1.800	8	1.842	8
510	1.730	11	1.771	13
521	1.611	7	1.648	8
440	1.560	43	1.596	37

Space group for both compounds is *P* $\bar{4}3n$.

Zn(NO₃)₂, 30 mmol H₃PO₄ and 20 mmol NaBr in ~30 cm³ H₂O and adding 84 mmol NaOH as a 2M solution. Initially this forms a gel, which transforms to a thick, creamy slurry on shaking, and settles, after standing overnight at pH 7.5, to a micro-crystalline sludge. After filtration, washing and drying, the powder diffraction pattern (Table 1) can be indexed cleanly using a primitive cubic structure with lattice parameter *a* = 8.824 Å. The powder is readily soluble in dilute acid, contains no bromide ion (and indeed NaBr is unnecessary in the synthesis), and loses about 12% of its mass as water on heating between 140 and 200 °C (the water content by mass for Na₆(H₂O)₈(ZnPO₄)₆ is 11.6%). At 200 °C the structure converts

TABLE 2 RHO analogues

<i>hkl</i>	Li ₂₄ Be ₂₄ As ₂₄ O ₉₆ · 40H ₂ O (<i>a</i> = 14.059 Å)		Rb ₂₄ Be ₂₄ As ₂₄ O ₉₆ · 40H ₂ O (<i>a</i> = 14.236 Å)	
	<i>D</i> _{obs}	Relative intensity	<i>D</i> _{obs}	Relative intensity
110	9.965	78	10.075	36
200	7.004	2	7.121	8
211	5.740	43	5.805	8
220	4.974	20	5.033	1
310	4.447	30	4.498	17
222	4.049	10	4.109	2
321	3.756	100	3.807	100
400	3.513	5	3.563	3
330	3.311	38	3.357	45
420	3.143	70	3.184	95
332	3.001	34	3.035	88
422	2.874	8	2.907	45
510	2.757	35	2.793	27
521	2.560	15	2.599	58
440	—	—	2.526	4
530	2.408	10	2.441	11
600	2.341	9	2.372	3
611	2.278	1	2.311	21
620	2.221	3	2.251	5
541	2.174	7	2.197	7
622	—	—	2.146	12
—	—	—	—	—

Space group for both compounds is *I*23.

TABLE 3 Zeolite X analogues

<i>hkl</i>	$\text{Na}_{96}\text{Be}_{96}\text{P}_{96}\text{O}_{192} \cdot 192\text{H}_2\text{O}$ ($a = 23.383 \text{ \AA}$)		$(\text{Na,TMA})_{96}\text{Zn}_{96}\text{P}_{96}\text{O}_{192} \cdot 192\text{H}_2\text{O}$ ($a = 25.226 \text{ \AA}$)	
	D_{obs}	Relative intensity	D_{obs}	Relative intensity
111	13.566	100	14.607	100
220	8.277	35	8.927	38
311	7.053	21	7.615	38
222	—	—	—	—
400	5.845	1	—	—
331	5.368	15	5.786	32
420	—	—	—	—
422	—	—	—	—
333	4.500	6	4.856	13
440	4.133	4	4.460	7
531	3.954	29	4.269	7
600	3.897	1	4.209	3
620	3.700	36	3.990	22
533	3.569	38	3.847	36
622	3.773	1	3.803	5
444	—	—	3.638	3
711	3.278	2	3.535	4
642	3.140	17	3.369	44
731	3.045	9	3.285	11
—	—	—	—	—

Space group for both compounds is $F\bar{d}3$.

smoothly to a hexagonal phase which, on heating to 700 °C, reverts to monoclinic NaZnPO_4 (refs 11, 12).

The corresponding arsenate is prepared by using arsenic acid instead of phosphoric acid (with or without halide ion), but it crystallizes somewhat more slowly, requiring a soak at 70 °C of two or three days. The microcrystalline powder can be indexed as a primitive cube with $a = 9.030 \text{ \AA}$ (Table 1). It loses 9.6% by mass of water on heating to 200–250 °C, converting to a hexagonal phase isotypic with the phosphate. This dehydrated phase can be rehydrated to the cubic sodalite with water (taking

48 h at 100 °C). At 550 °C, the material converts to NaZnAsO_4 (ref. 12).

RHO analogues. $\text{Li}_{24}\text{Be}_{24}\text{P}_{24}\text{O}_{96} \cdot 40\text{H}_2\text{O}$ has been prepared by Harvey and Meier⁹. We have modified their preparation slightly by eliminating the use of tetraethyl ammonium hydroxide and operating below 100 °C, because at this temperature the RHO phase transforms to the Li-A(BW) structure (see below). The arsenate analogue (DPZ-2B) has been prepared as follows. 12 mmol $\text{Be}(\text{NO}_3)_2$ and 14 mmol H_3AsO_4 were dissolved in 15 cm³ of water, and after thorough mixing, 38 mmol LiOH in 10 cm³ of water was added. The resultant gel was shaken vigorously, upon which it transformed to a creamy milk (pH 7.5). After standing for 8 h at 70 °C, the settled microcrystalline slurry was recovered by filtration and air drying. The powder exhibited a clean body-centred cubic X-ray diffraction pattern with lattice parameter $a = 14.059 \text{ \AA}$ (Table 2). A thermogravimetric analysis indicated a mass loss of 16.5% as water on heating to 300 °C (water content by mass = 16.23% for $\text{Li}_{24}\text{Be}_{24}\text{As}_{24}\text{O}_{96} \cdot 40\text{H}_2\text{O}$). Some degradation of the ideal structure accompanied water loss.

Another RHO structure (DPZ-2A) with the same beryllo-arsenate framework has been made by using RbOH instead of LiOH and soaking the initial gel-slurry at 70 °C for two days. The X-ray powder diffraction pattern could be cleanly indexed as body-centred cubic, $a = 14.236 \text{ \AA}$ (Table 2). The material seems to be considerably more stable than the lithium analogue, losing water only slowly (9% mass loss at 600 °C; calculated water content for $\text{Rb}_{24}\text{Be}_{24}\text{As}_{24}\text{O}_{96} \cdot 40\text{H}_2\text{O} = 11.39\%$) with full retention of structure. The use of mixtures of LiOH and RbOH at templates yields materials with cubic cell dimensions between those of the endmembers, but Na and K analogues cannot be prepared in this way.

Zeolite X analogues. Sodium beryllophosphate X, $\text{Na}_{96}\text{Be}_{96}\text{P}_{96}\text{O}_{192} \cdot 192\text{H}_2\text{O}$ (DPZ-1A) is made most conveniently by preparing a clear solution of 235 mmol NaOH and 84 mmol H_3PO_4 in 80 cm³ of water and adding a solution of 72 mmol $\text{Be}(\text{NO}_3)_2$ in about 35 cm³ of water. On shaking, the initial gel converts to a milk which settles completely on standing for 12 h at 70 °C. The diffraction pattern of the recovered material can be indexed very well as face-centred cubic, $a = 23.383 \text{ \AA}$ (Table

TABLE 4 Zeolite Li-A(BW) analogues

$\text{Li}_4\text{Be}_4\text{P}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ $a = 7.817 \text{ \AA}$ $b = 9.662 \text{ \AA}$ $c = 4.739 \text{ \AA}$ $V = 357.9 \text{ \AA}^3$			$\text{Li}_4\text{Be}_4\text{As}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ $a = 8.018 \text{ \AA}$ $b = 10.043 \text{ \AA}$ $c = 4.854 \text{ \AA}$ $V = 390.9 \text{ \AA}^3$			$\text{Li}_4\text{Zn}_4\text{P}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ $a = 8.122 \text{ \AA}$ $b = 10.492 \text{ \AA}$ $c = 4.854 \text{ \AA}$ $V = 428.0 \text{ \AA}^3$			$\text{Li}_4\text{Zn}_4\text{As}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ $a = 8.289 \text{ \AA}$ $b = 10.816 \text{ \AA}$ $c = 5.156 \text{ \AA}$ $V = 462.2 \text{ \AA}^3$		
hkl	D_{obs}	Relative intensity	D_{obs}	Relative intensity	D_{obs}	Relative intensity	D_{obs}	Relative intensity	D_{obs}	Relative intensity	
110	6.075	82	6.269	99	6.423	100	6.581	100			
020	4.828	21	5.024	19	5.246	9	5.408	18			
120	4.108	6	4.255	19	4.408	21	4.532	8			
101	4.049	47	4.151	50	4.270	43	4.378	32			
200	3.902	4	4.011	6	4.056	5	4.144	6			
111	3.737	27	3.837	60	3.996	3	4.059	10			
210	3.621	24	3.724	45	3.788	3	3.869	6			
021	3.383	24	3.490	46	3.627	3	—	—			
121	3.110	6	3.200	6	—	—	—	—			
130	2.979	100	3.089	100	3.214	49	3.307	55			
211	2.877	69	2.955	85	3.023	64	3.095	41			
221	—	—	2.632	5	—	—	—	—			
131	—	—	—	—	2.706	6	2.779	6			
310	2.516	4	—	—	—	—	—	—			
230	—	—	—	—	2.650	10	2.719	23			
002	2.369	25	2.427	24	2.510	21	2.578	12			
301	2.281	7	2.341	13	2.384	8	2.435	8			
112	—	—	—	—	2.337	8	—	—			
231	—	—	2.271	12	—	—	2.405	10			
400	1.954	3	2.005	5	2.030	11	2.072	7			

Space group for all compounds is $Pna2_1$.

3). The powder loses 22% mass as water gradually on heating to 350 °C (calculated water content = 22.1%) with marginal retention of structure. Further heating to 700 °C results in conversion to monoclinic beryllonite, NaBePO_4 , with no weight loss. If the initial synthesis mixture is heated for too long at 70 °C (or more rapidly at 100 °C), the product converts to Harvey and Meier's beryllphosphate G, a gismondine derivative.

Another zeolite X analogue (DPZ-1B) is available in the zincophosphate system, using sodium and tetramethyl ammonium ions as templates. A clear solution, containing 32 mmol NaOH, 134 mmol tetramethyl ammonium hydroxide and 64 mmol H_3PO_4 , was prepared in 75 cm³ of water and cooled to about 4 °C. To this was added a pre-cooled solution of 48 mmol $\text{Zn}(\text{NO}_3)_2$ to give a gel that converts to a milk on shaking. The milk settles rapidly and the solid may be recovered after standing for 5 h at 4 °C. The X-ray powder diffraction pattern is indexed as face-centred cubic, $a = 25.226 \text{ \AA}$. The product seems to be fairly sensitive to the ratio of sodium to zinc in the synthesis mixture: if less than 2/3, hopeite ($\text{Zn}_3(\text{PO}_4)_2 \cdot 4\text{H}_2\text{O}$) is formed, and if greater, a variety of hydrated sodium zinc phosphates (including the sodalite phase described above) contaminate the product. It is tempting to speculate that the sodium ion acts as a template for 'sodalite' cubes in solution which then condense around the large alkyl ammonium ions to form the X structure.

Zeolite Li-A(BW) analogues. Four new materials with the lithium aluminosilicate structure ($\text{Li}_4\text{Al}_4\text{Si}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ (orthorhombic $\text{Pna}2_1$)¹³ have been made in which Be or Zn completely replace Al, and P or As replace Si.

(1) $\text{Li}_4\text{Be}_4\text{P}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ (DPZ-4A). This analogue often begins to appear if the RHO synthesis described above is carried out at higher temperature (100 °C), and may be the impurity phase noticed by Harvey and Meier⁹. The synthesis is optimized by using 12 mmol $\text{Be}(\text{NO}_3)_2$ and 14 mmol H_3PO_4 in 20 cm³ of water with subsequent addition of 37 mmol LiOH in 9 cm³ of water (pH 2.5). After soaking overnight at 100 °C, all RHO lines have disappeared from the diffraction pattern, leaving the pure Li-A(BW) phase (orthorhombic, Table 4). The phase is unstable to water loss (13% by mass, occurring in three steps on heating to 350 °C; calculated water content = 13.96%) with loss of the initial structure.

(2) $\text{Li}_4\text{Be}_4\text{As}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ (DPZ-4B). Carrying out the equivalent beryllarsenate RHO synthesis at lower pH (37 mmol LiOH, pH = 3) and higher temperature (100 °C) results in rapid (overnight) disappearance of the initially formed RHO phase, and conversion to the Li-A(BW) material (orthorhombic, Table 4). Water loss is complete at 260 °C (two steps, 10% by mass; calculated value = 10.4%).

(3) $\text{Li}_4\text{Zn}_4\text{P}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ (DPZ-4C) and (4) $\text{Li}_4\text{Zn}_4\text{As}_4\text{O}_{16} \cdot 4\text{H}_2\text{O}$ (DPZ-4D). These Li-A(BW) analogues are easily prepared in a similar way to the Be counterparts, but lower temperatures suffice (phosphate: 36 h at 70 °C, pH 7; arsenate: 48 h at 25 °C, pH 7). The arsenate in particular recrystallizes readily to the phenakite analogue¹⁴ LiZnAsO_4 if the synthesis is carried out at 100 °C. The powder diffraction patterns for all of these phases are detailed in Table 4.

The pH, water concentration, heating time, ionic ratios and solution homogeneity before gel formation all play a critical role in the synthesis of these phases. For example, we have obtained seven different hydrated sodium zinc phosphates by simple variation of these synthesis conditions. Further chemical and structural characterization of these and other materials is in process and will be published in more detail elsewhere (T.E.G., G.D.S. *et al.*, manuscript in preparation). □

- Wang, Tieli *et al.* *J. Chem. Soc. Chem. Comm.* **14**, 948-949 (1989).
- Palache, C., Berman, H. & Frondel, C. (eds) *Dana's System of Mineralogy* Vol. 2, 997 (Wiley, New York, 1951).
- Peacor, D. R., Rouse, R. C. & Ahn, J.-H. *Am. Miner.* **72**, 816-820 (1987).
- Rouse, R. C. *Jb. Miner. Mh.* **10**, 433-440 (1987).
- Harvey, G. & Meier, W. M. in *Studies in Surf. Science & Catalysis* **49A** (eds Jacobs, P. A. & VanSanten, R. A.) 411-420 (Elsevier, Amsterdam, 1989).
- Felsche, J., Luger, S. & Baerlocher, Ch. *Zeolites* **6**, 367-372 (1986).
- Elammari, L., Durand, J., Cot, L. & Elouadi, B. *Z. Kristallogr.* **180**, 137-140 (1987).
- Andratschke, M. & Feltz, A. *Z. Chem.* **28**, 112-113 (1988).
- Kerr, I. S. *Z. Kristallogr.* **139**, 186-195 (1974).
- Blasse, G. *J. inorg. nucl. Chem.* **25**, 136 (1963).

ACKNOWLEDGEMENTS. This research was supported in part by an NSF grant (G.D.S.).

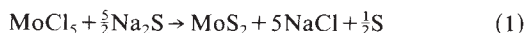
Rapid solid-state synthesis of materials from molybdenum disulphide to refractories

Philippe R. Bonneau, Robert F. Jarvis Jr & Richard B. Kaner*

Department of Chemistry and Biochemistry and the Solid State Science Center, University of California, Los Angeles, California 90024-1569, USA

LAYERED molybdenum and tungsten dichalcogenides have stimulated considerable interest as lubricants^{1,2}, battery cathodes³⁻⁵, and catalysts⁶⁻⁸. Pure, crystalline group VI transition-metal dichalcogenides are normally prepared by intermittently grinding and heating the elements at >900 °C for several days. Low-temperature solution routes to these materials have also been explored⁹⁻¹¹. In those studies, exchange (metathesis) reactions between transition-metal halides and alkali-metal sulphides or covalent sulphiding agents in polar organic solvents were found generally to yield finely divided products at close to ambient conditions. Here, in contrast, we consider the factors that influence reactions between transition-metal halides and alkali-metal chalcogenides in the solid state. These highly energetic reactions, driven by the formation of very stable product species, provide a powerful method for the rapid synthesis of materials normally prepared at high temperatures over long periods of time. Other advantages of these solid-state reactions include control of reaction conditions and of product particle sizes. Although our study focuses on group VI transition-metal dichalcogenides, analogous metathesis reactions can be used for rapid syntheses, initiated at low temperatures, of many other technologically important materials.

Reactions between transition-metal (V, VI) chlorides and alkali-metal sulphides are extremely vigorous, highly exothermic and visually dramatic. For example, the redox reaction



can self-ignite or be detonated with a hot filament (at 40-60 °C), producing an intense flash of light and generating molten sodium chloride. Some of the more volatile reaction components, such as excess sulphur and any unreacted metal chloride, are volatilized instantly. Crystallizing within a few seconds in a rapidly cooling salt flux, pure molybdenum disulphide is then isolated simply by washing with chloroform or carbon disulphide, water and methanol.

The crystallinity of the product is remarkable, especially considering the reaction conditions and the reaction time. Figure 1 shows the powder X-ray diffraction patterns of MoS_2 prepared from the elements heated at 900 °C for five days in an evacuated silica tube (a) and synthesized using the rapid, solid-state precursor method (b). The crystallinities of the two different samples are similar. The metathesis reaction yields both the 2H

* To whom correspondence should be addressed.

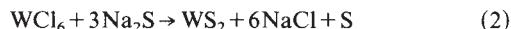
Received 19 October; accepted 10 December 1990.

- Breck, D. W. *Zeolite Molecular Sieves* (Wiley, New York, 1974).
- Barrer, R. M. *Zeolite and Clay Minerals* (Academic Press, London, 1978).
- Szostak, R. *Molecular Sieves* (Van Nostrand Reinhold, New York, 1989).
- Mooney-Slater, R. C. L. *Acta Crystallogr.* **20**, 526-534 (1966).

and 3R polytypes of MoS_2 , whereas the high-temperature reaction of the elements yields only the 2H polytype. Both diffraction patterns contain sharp 00 l and $hk0$ lines, indicating, respectively, uniform stacking of layers and long-range order within sheets. Broadened $h0l$ lines are a consequence of layer rotations about the axis perpendicular to the basal planes, resulting in a lack of consistent orientation between sheets¹².

The large amount of heat generated in the precursor reaction is responsible for the high degree of crystallinity of the product. Bomb calorimetry yields an average value of $-174 \pm 5 \text{ kcal mol}^{-1}$ for the reaction between molybdenum pentachloride and sodium monosulphide. The calculated enthalpy of reaction based on Hess' Law is $-213 \text{ kcal mol}^{-1}$ (ref. 13). The experimental value, $\sim 82\%$ of the theoretical value, is therefore in good agreement with an 80% reaction yield. Optical pyrometry reveals that the molybdenum disulphide reaction reaches $\sim 1,050^\circ\text{C}$ within 1 s of initiation. This value was obtained for 1.714 g of reactant mixture and does not take into account the effects of a non-uniform radiation source and the heat capacity of the container.

An analogous reaction, in which tungsten hexachloride replaces molybdenum pentachloride, produces crystalline tungsten disulphide.



This reaction is also quite exothermic, producing $-188 \pm 7 \text{ kcal mol}^{-1}$. The calculated enthalpy of reaction in this case is $-252 \text{ kcal mol}^{-1}$ (ref. 13).

These precursor reactions are relatively straightforward and easily controllable provided that the reagents are carefully pre-

pared. The sulphide reagent must be sintered to decrease its surface area. Sodium monosulphide, freshly precipitated from liquid ammonia at -30°C , has a very small particle size and a correspondingly high surface area. Mixing the unsintered sulphiding agent with either of the transition-metal (V or VI) chlorides results in appreciable surface reaction. The heat generated by the surface reaction is, in most cases, sufficient to initiate a bulk reaction, and the mixture self-ignites. If, however, the sulphiding agent is sintered at $500\text{--}600^\circ\text{C}$ for one day to decrease its surface area, the reagents may then be combined without fear of self-initiated reaction. Sintered sulphiding agents were needed in order to carry out bomb calorimetry successfully.

The transition-metal chlorides are resublimed to remove any lower-valent and less volatile metal chlorides. If the latter are not separated from the molecular precursors, the reactions, although still spontaneous, are less vigorous, products are less crystalline, and yields are noticeably lower. Solid-state reactions between transition-metal (IV, III, and II) chloride precursors and sodium monosulphide, however, are not as easily activated. Even at $200\text{--}250^\circ\text{C}$, only surface reactions occur and the bulk of the reagents remain unreacted. Many of these reactions take place at $\sim 300^\circ\text{C}$, at which point many of the transition-metal chlorides either sublime or decompose to initiate the reaction.

The choice of the transition-metal halide precursor is therefore critical. It affects the spontaneity and the yield of the reaction as well as the crystallinity of the transition-metal dichalcogenide product. These effects may be considered from various perspectives, but all are related to the oxidation state of the transition metal in the precursor and its structure.

First, examination of the structures of several halide precursors reveals that lower-valent halides are normally three-dimensional salts, infinite sheets, or polymeric chains; hexavalent and pentavalent halides, however, are usually monomers, dimers or tetramers¹⁴. The molecular species are much more volatile than the transition-metal halide salts with network structures. Considerably less energy is required to break down a molecular solid than is needed to break down an ionic lattice. More heat is therefore available to crystallize the product. The highly volatile monomers and dimers are also more readily available for reaction with the sulphiding agent. Thus solid-state metathesis reactions involving molybdenum pentachloride or tungsten hexachloride are initiated at lower temperatures and produce more-crystalline products.

Second, solid-solid reactions are generally exothermic, and the driving force for these reactions is the difference between the free energies of formation of the products and the reactants. The thermodynamics of these reactions indicate that the main driving force is not the formation of the transition-metal dichalcogenide product but the formation of sodium chloride. Molecular species such as molybdenum pentachloride and tungsten hexachloride contain a higher ratio of halide atoms to transition-metal atoms. Stoichiometric reactions between a molecular transition-metal halide precursor and an alkali-metal sulphiding agent therefore produce more salt by-product and, consequently, more heat to crystallize the product(s). The enthalpies for reactions between MoCl_5 , MoCl_4 , or MoCl_3 with Na_2S are -213 , -171 and $-132 \text{ kcal mol}^{-1}$ respectively¹³. Therefore, the crystallinity of the transition-metal dichalcogenide product depends on the choice of the metal halide precursor.

Finally, both the reduction of the transition-metal cation and the oxidation of the sulphide anion in the above metathesis reactions are spontaneous processes. It is reasonable to expect that a greater difference in oxidation state between the transition metal and the sulphur in the precursor species provides a greater driving force in these redox reactions. Reactions containing transition-metal (V, VI) halides should be, and indeed are, more spontaneous than reactions containing transition-metal (IV, III, or II) halides.

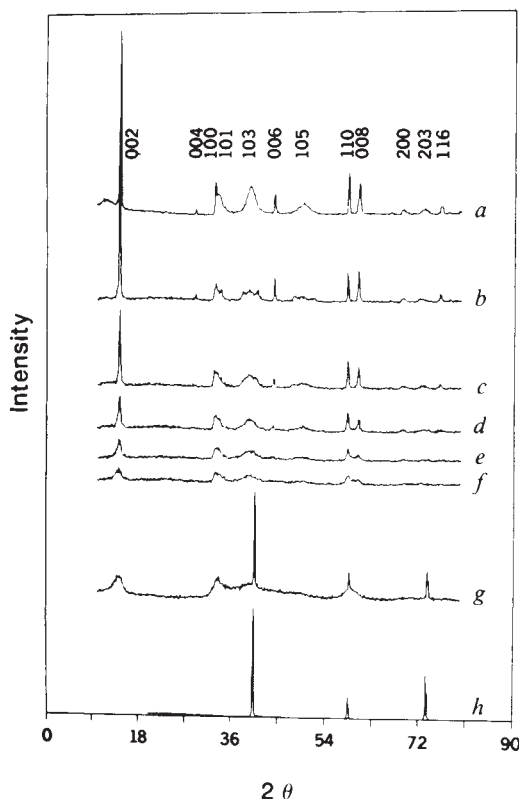


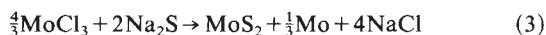
FIG. 1 X-ray diffraction patterns of MoS_2 produced (a) from the elements heated to 900°C for 5 days and (b–f) in $<5 \text{ s}$ from the reaction of MoCl_5 and Na_2S . In c–f, NaCl was added in the following ratios to 1.714 g of reactant mixture: c, 0.25:1; d, 0.5:1; e, 1:1; and f, 2:1. MoS_2 and Mo metal formed in the reaction between MoCl_3 and Na_2S at 300°C for four days (g). h, Diffraction pattern for Mo metal. X-ray diffraction scans were obtained using Ni-filtered $\text{Cu K}\alpha$ radiation at increments of $0.05^\circ 2\theta$ at 5 s per increment.

TABLE 1 Thermogravimetric analyses and X-ray diffraction results

Compound	% weight of metal		Calculated Formula	Hexagonal lattice parameters (Å)
	Experimental	Theoretical		
MoS ₂	60.1	59.9	MoS _{1.99}	<i>a</i> = 3.163 <i>c</i> = 12.29
MoSe ₂	38.0	37.8	MoSe _{1.98}	<i>a</i> = 3.290 <i>c</i> = 12.93
WS ₂	73.9	74.1	WS _{2.02}	<i>a</i> = 3.161 <i>c</i> = 12.36
WSe ₂	54.0	53.8	WSe _{1.98}	<i>a</i> = 3.288 <i>c</i> = 13.00

Thermogravimetric analyses were carried out in a stream of 5% H₂/95% N₂. The temperature was increased by 20 °C per minute to 900 °C. The sample remained under these conditions until no further weight loss was recorded (24–40 h). Powder X-ray diffraction on all TGA residues revealed only pure Mo or W. Lattice parameters were determined by a least-squares fit of 00*l* and *h*k0 lines occurring between 10° and 80° 2θ.

Results of the reaction of molybdenum trichloride with sodium monosulphide illustrate the above points. In addition, an interesting feature of this reaction is that the sodium-to-chloride and molybdenum-to-sulphide ratios are never stoichiometric with respect to the formation of the transition-metal dichalcogenide. At the higher temperatures necessary for reaction, molybdenum disulphide is still the favoured phase, even under nonstoichiometric conditions. The net reaction between molybdenum trichloride and sodium monosulphide is then:



Powder X-ray diffraction of the residue of reaction (3) confirms the presence of these products. Reactions of molybdenum trichloride with sodium monosulphide indeed lead to the formation of much-less-crystalline molybdenum disulphide and appreciable amounts of molybdenum (Fig. 1g). The presence of Mo metal is undesirable, as it is not easily separated from the molybdenum dichalcogenide product. The results of varying the oxidation state of the transition metal in the halide precursor are summarized in Fig. 1b and g. In general, the crystallinity of the product decreases as the oxidation state of the metal in the precursor decreases.

A study with nickel sulphides revealed that transition-metal dichalcogenide products may be strongly influenced by the sulphiding agent¹⁵. When more than one stoichiometric phase is thermodynamically favourable, the chalcogenide-rich phase is formed preferentially in the presence of a sulphur-rich alkali-metal sulphiding agent, such as Na₂S₂, Na₂S₄ and Na₂S₅. Reactions between molybdenum pentachloride or tungsten hexachloride and sodium monosulphide do not, however, lead to mixed stoichiometric phases: molybdenum and tungsten disulphides are favoured overwhelmingly under these reaction conditions.

To determine the effect on the reaction and products of sulphur-rich sulphiding agents, we carried out reactions between molybdenum pentachloride and Na₂S₂ or Na₂S₅. An increased sulphur content leads to a decrease in the crystallinity of the dichalcogenide product. This phenomenon can be understood in terms of the total heat available relative to the total mass in

the system. Substitution of Na₂S₂ or Na₂S₅ for Na₂S does not change the enthalpy of reaction appreciably, but it does contribute a large amount of excess sulphur. A considerable amount of the heat released during the reaction is absorbed by the more volatile components. Species not immediately contributing to the reaction serve only to draw from the system heat that is necessary for crystallization of the transition-metal dichalcogenide. Extraneous materials act as a heat sink, decreasing local heating and favouring a less crystalline product¹⁶.

One should therefore be able to introduce inert compounds such as sodium chloride, a by-product of these reactions, to effectively control the crystallinity of the product. Figure 1b–f shows the results of a series of experiments designed to illustrate this point. A fixed amount (1.714 g) of reactant mixture is combined with sodium chloride in the indicated weight ratios. The addition of increasingly larger non-reactive masses evidently leads to smaller product particle sizes, reflected in the progressively broadened lines of the powder X-ray diffraction patterns. Furthermore, when ratios smaller than 1:2 of reactants to added sodium chloride are used, the bulk of the mixture can no longer be ignited with a heated filament. Large amounts of inert material may therefore be used to control the reaction conditions as well as the product particle sizes.

Solvents have been used as reaction media and heat sinks to control crystallite sizes in a number of systems⁹. Unfortunately, many desirable reagents such as alkali-metal selenides, nitrides, phosphides and arsenides are solvent incompatible (for example, Li₃N, Na₂Se, Na₃P and Na₃As). Some are virtually insoluble, and others decompose in or react (sometimes violently) with many solvents. In addition, certain metathesis reactions require initiation temperatures that would readily decompose most solvents. Solid-state metathesis reactions do not have these limitations, and therefore they provide a viable synthetic route for an extensive range of materials. Moreover, alkali-metal pnictide and chalcogenide reagents are preferable to their gaseous, highly toxic counterparts H₂S, H₂Se, PH₃ and AsH₃, because of the ease and safety of handling.

These solid-state metathesis reactions are attractive because of their applicability to other metal and anion systems. Molybdenum and tungsten diselenides are prepared simply by replacing sodium monosulphide with sodium monoselenide. Thermogravimetric analyses and powder X-ray diffraction data indicate that the transition-metal diselenide products are phase-pure and crystalline (Table 1). Tantalum and niobium-pentachlorides, like molybdenum pentachloride and tungsten hexachloride, react with sodium monoselenide to form the corresponding transition-metal diselenides. Reactions of the above transition-metal halides with sodium monophosphide yield transition-metal monophosphides. Zirconium nitride, a refractory ceramic (melting point ~3,000 °C), can be synthesized using zirconium tetrachloride and lithium nitride. Gallium arsenide, a III–V semiconductor, can be prepared by reacting gallium triiodide and sodium arsenide. Mixed-transition-metal and mixed-anion solid solutions such as (Mo, W)S₂, Mo(S, Se)₂ and Sm(P, As) are accessible via the solid-solution precursors MoCl₅ · WCl₆, Na₂(S, Se) and Na₃(P, As), respectively. We are now studying the application of rapid solid-state syntheses to other technologically important compounds. □

Received 24 September; accepted 11 December 1990.

1. Clauss, F. L. *Solid Lubricants and Self-Lubricating Solids* (Academic, New York, 1972).
2. Lince, J. R. & Fleischauer, P. D. *J. Mater. Res.* **2**, 827–838 (1987).
3. Rouxel, J. & Brec, R. A. *Rev. Mater. Sci.* **16**, 137–162 (1986).
4. Auburn, J. J., Barberio, Y. L., Hansen, K. J., Schleich, D. M. & Martin, M. J. *J. electrochem. Soc.* **134**, 580–586 (1987).
5. Jacobson, A. J., Chianelli, R. R. & Whittingham, M. S. *J. electrochem. Soc.* **126**, 2277–2278 (1979).
6. Chianelli, R. R. *Catal. Rev. Sci. Engng* **26**, 361–393 (1984).
7. Pecoraro, T. A. & Chianelli, R. R. *J. Catal.* **67**, 430–445 (1981).
8. Gobolos, S., Wu, Q., Delannay, F., Grange, P. & Delmon, B. *Polyhedron* **5**, 219–224 (1986).
9. Chianelli, R. R. & Dines, M. B. *Inorg. Chem.* **17**, 2758–2762 (1978).

10. Schleich, D. M. & Martin, M. J. *J. Solid St. Chem.* **64**, 359–364 (1986).
11. Martin, M. J., Qiang, G. & Schleich, D. M. *Inorg. Chem.* **27**, 2804–2808 (1988).
12. Chianelli, R. R. *Int. Rev. Phys. Chem.* **2**, 127–165 (1982).
13. JANAF Thermochemical Tables, 3rd Edn (ed. Lide, D. R. Jr) (American Chemical Society and American Institute of Physics, Inc., New York, 1985).
14. Wells, A. F. *Structural Inorganic Chemistry* 5th Edn (Clarendon, Oxford, 1986).
15. Bonneau, P. R., Shibao, R. K. & Kaner, R. B. *Inorg. Chem.* **29**, 2511–2514 (1990).
16. Chakurov, Chr., Rusanov, V. & Koichev, J. *J. Solid St. Chem.* **71**, 522–529 (1987).

ACKNOWLEDGEMENTS. This work was supported by the NSF through the Presidential Young Investigator Award Program and by a David and Lucille Packard Foundation Fellowship in Science and Engineering.

Initiation of Fennoscandian ice-sheet retreat during the last deglaciation

S. J. Lehman*, G. A. Jones*, L. D. Keigwin*,
E. S. Andersen†, G. Butenko‡ & S-R. Østmo‡

* Woods Hole Oceanographic Institution, Woods Hole,
Massachusetts 02543, USA

† University of Oslo, Blindern N-0316, Oslo 3, Norway

‡ Norsk Hydro A/S, N-1321 Stabekk, Norway

ALTHOUGH the retreat history of the southern margin of the Laurentide ice sheet during the last deglaciation has been well known for several decades¹⁻³ and recently supported by evidence from accelerator mass spectrometer (AMS) radiocarbon dating of meltwater in the Gulf of Mexico⁴ and the North Atlantic subtropical gyre⁵, the retreat history of the Fennoscandian ice sheet before 13 kyr ago is still poorly documented. From AMS ¹⁴C dating and isotopic analysis of a rapidly deposited series of supratill sediments in the Norwegian Channel (North Sea), we find that the southern margin of the Fennoscandian ice sheet was in retreat by 15 kyr BP (before present), approximately 2,000 yr earlier than previously supposed. Oxygen isotope analyses in the channel sediments confirm earlier evidence from deep Norwegian Sea sediment cores for low salinity due to ice sheet discharge beginning ~15–14.5 kyr BP (ref. 6). Recent recalibration of the ¹⁴C timescale

indicates that the onset of deglaciation within the Norwegian Sea occurred during an interval of rapidly increasing summer insolation beginning ~18,000 calendar yr BP (ref. 7), suggesting that ice-sheet retreat may have been caused by increasing insolation acting on a particularly climate-sensitive ice-sheet configuration. Fresh water and icebergs discharged into the Norwegian Sea at this time reached the North Atlantic and may have contributed to a brief interval of extreme oceanic cooling⁸ and reduced production of North Atlantic Deep Water^{9,10}. These factors may also have helped to briefly reverse the deglaciation trend (Erie Interstade) already underway in North America.

The former western and southwestern margins of the Fennoscandian ice sheet extended onto the continental shelf during the Last Glacial Maximum (LGM; ~18 kyr BP), and it has been widely supposed, on the basis of off-shore mollusc ¹⁴C dates that deglaciation of the shelf occurred ~13 kyr BP¹¹. But in Denmark, a few reliable mollusc ¹⁴C dates as old as 14 kyr indicate that the ice had retreated ~50 km from the maximum position by that time¹². New AMS ¹⁴C dates on gyttja and mosses from Southwest Norway suggest deglaciation of the adjacent shelf by 14 kyr BP, but these may be erroneously old owing to hard-water effects and contamination by reworked carbon^{13,14}. Thus, a secure chronology of Fennoscandian deglaciation before 13 kyr is clearly needed.

We have studied the upper 25 m of core Troll 3.1 (60°46.7' N, 3°42.8' E), a 95-m-long Shelby tube sediment core (>95% recovery) retrieved from the central part of the Norwegian Channel at a water depth of 332 metres. The core site (Fig. 1) lies ~80 km inside the proposed minimum LGM ice-sheet recon-

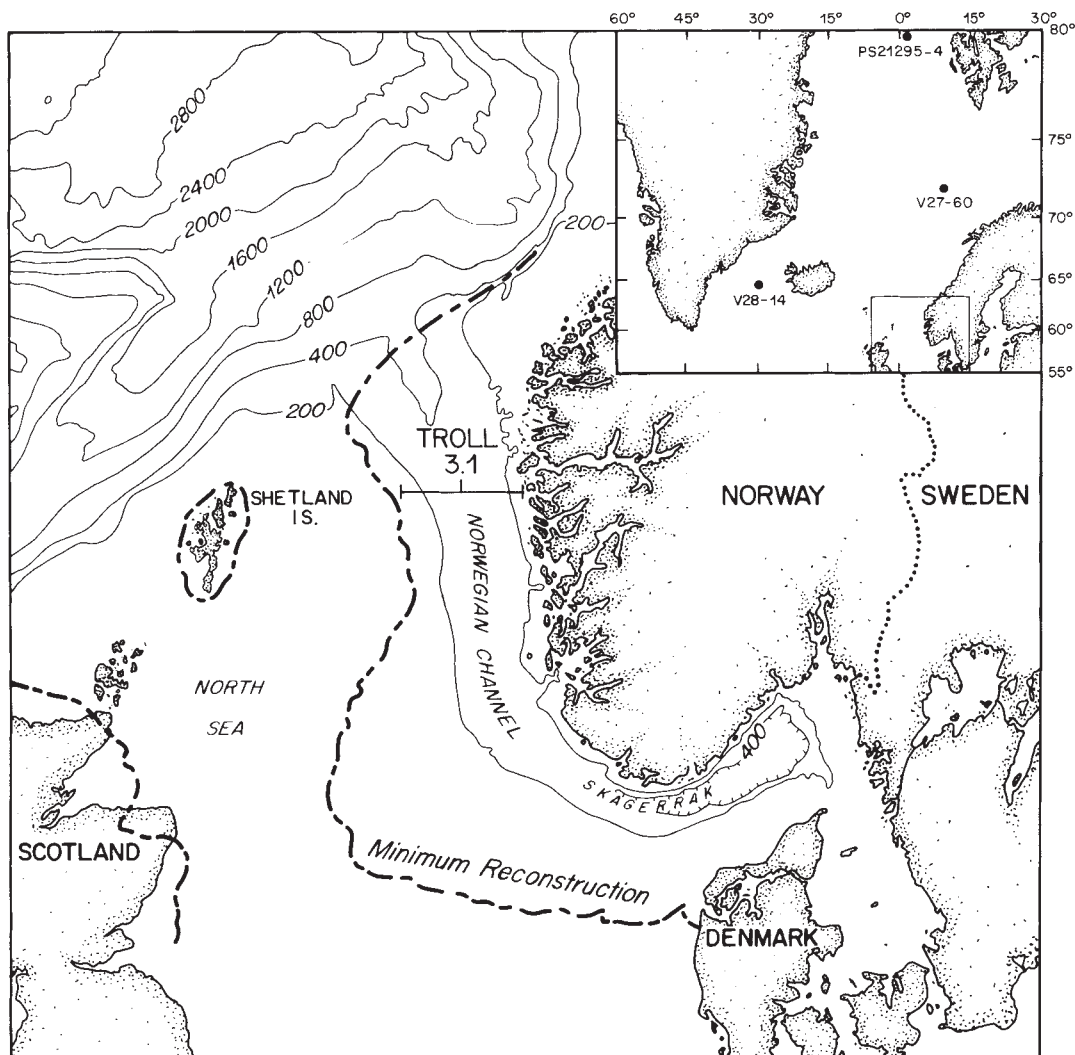


FIG. 1 Location of core Troll 3.1 in the Norwegian Channel with respect to the minimum LGM ice-sheet reconstruction (refs 15, 16) and the generalized geological section given in Fig. 2.

TABLE 1 AMS ^{14}C dates of benthic foraminifera

Depth (m)	Species	$\delta^{18}\text{O}(\text{‰})$ PDB	corresponding ^{14}C age (yr)	Target no.	Accession no.	Dated species
0.175–0.22	<i>C. laev</i>	2.23	$1,490 \pm 60$	WHG692	AA5324	<i>Uvigerina peregrina</i>
0.58–0.62	<i>C. laev</i>	2.32				
0.78–0.82	<i>C. laev</i>	2.31				
1.28–1.32	<i>C. laev</i>	2.11				
1.48–1.52	<i>C. laev</i>	2.30				
1.68–1.72	<i>C. laev</i>	2.14	$7,060 \pm 90$	WHG694	AA5380	<i>U. per</i>
1.88–1.92	<i>C. laev</i>	2.12				
2.28–2.32	<i>C. laev</i>	2.18				
2.48–2.52	<i>C. laev</i>	2.43				
2.88–2.92	<i>C. laev</i>	2.75				
3.30–3.40	<i>C. laev</i>	2.91	$10,510 \pm 120$	WHG695	AA5326	<i>Nonion labradoricum</i>
3.70–3.80	<i>C. laev</i>	3.02				
4.30–4.40	<i>C. laev</i>	3.66				
4.30–4.40	<i>N. lab</i>	3.87				
4.50–4.60	<i>N. lab</i>	4.04				
4.70–4.80	<i>N. lab</i>	3.80	$11,160 \pm 120$	WHG880	AA5946	<i>N. lab</i>
4.90–4.98	<i>N. lab</i>	3.87				
5.30–5.40	<i>N. lab</i>	3.73				
5.50–5.60	<i>N. lab</i>	3.77				
5.70–5.80	<i>N. lab</i>	3.79				
5.85–5.93	<i>N. lab</i>	3.82	$11,730 \pm 100$	WHG696	AA5381	<i>N. lab</i>
6.40–6.50	<i>N. lab</i>	4.38				
6.50–6.60	<i>N. lab</i>	4.33				
6.70–6.80	<i>N. lab</i>	3.96				
6.86–6.94	<i>N. lab</i>	4.12				
7.30–7.40	<i>N. lab</i>	4.19	$12,410 \pm 110$	WHG697	AA5382	<i>N. lab</i>
7.50–7.60	<i>N. lab</i>	4.37				
7.70–7.80	<i>N. lab</i>	4.42				
7.89–7.97	<i>N. lab</i>	4.45				
8.50–8.60	<i>N. lab</i>	4.52				
8.70–8.80	<i>N. lab</i>	4.39	$12,590 \pm 150$	WHG699	AA5383	<i>N. lab</i>
8.89–8.97	<i>N. lab</i>	4.24				
9.29–9.40	<i>N. lab</i>	4.25				
9.50–9.60	<i>N. lab</i>	4.33				
9.70–9.80	<i>N. lab</i>	4.21				
9.85–9.93	<i>N. lab</i>	4.10	$13,450 \pm 160$	WHG698	AA5327	<i>N. lab</i>
10.47–10.60	<i>N. lab</i>	3.99				
11.25–11.35	<i>N. lab</i>	3.97				
11.83–11.93	<i>N. lab</i>	3.85				
13.80–13.90	<i>N. lab</i>	4.09	$14,690 \pm 120$	WHG701	AA5329	<i>N. lab</i>
14.15–14.25	<i>N. lab</i>	4.27				
14.74–14.84	<i>N. lab</i>	4.32				
15.40–15.50	<i>N. lab</i>	4.29				
15.85–15.95	<i>N. lab</i>	4.30				
17.47–17.57	<i>N. lab</i>	4.19	$>49,000$	WHG881	AA5947	<i>Mixed benthics</i>
18.82–18.92	<i>N. lab</i>	3.88				
20.48–20.58	<i>N. lab</i>	3.33				
20.85–20.95	<i>N. lab</i>	3.43				
21.19–21.29	<i>N. lab</i>	3.42				
21.88–21.98	<i>N. lab</i>	3.48	$>49,000$	WHG881	AA5947	<i>Mixed benthics</i>
22.19–22.29	<i>N. lab</i>	3.38				
22.88–22.98	<i>N. lab</i>	3.64				
25.20–25.30	—	—	$>49,000$	WHG881	AA5947	<i>Mixed benthics</i>

struction for the area^{15,16} and along the main axis of the proposed connection between the Fennoscandian and British ice sheets in the maximum LGM reconstruction¹⁷. Because the site is located within a trough, it is ideally placed to record the first marine incursion to the area signifying ice-sheet retreat. A generalized cross-section of the late Quaternary stratigraphy in the channel, from high-resolution multichannel seismic reflection profiles and borehole studies^{18–20}, shows a prominent subsurface reflector (reflector IIa of Fig. 2) that can be traced from the coast of Norway, across the channel, and onto the edge of the North Sea Plateau. This reflector has previously been interpreted as a glacial surface. Lithological and geotechnical studies^{18,21} have shown that sediments below the 25-m level in Troll 3.1 have decreased water content, increased shear strength, high percentages of detrital carbonate and high stone content, confirming that the sediment was subglacially deposited and

that the 25-m level in Troll 3.1 is correlated with the regional glacial reflector. A conventional ^{14}C date of $18,860 \pm 260$ yr BP on derived shell fragments from beneath the glacial reflector on the North Sea Plateau suggests that the channel was glaciated at least once since that time¹⁹, for example, during the LGM (~ 18 kyr BP).

AMS ^{14}C dates were obtained from a series of nine monospecific samples of benthic foraminifera separated from glacial marine and normal marine sediments overlying reflector IIa in Troll 3.1 (Table 1; Fig. 3). The deepest sample above the reflector containing sufficient foraminifera for monospecific dating, at 22.83–22.93 m below seabed, yielded an age of $14,690 \pm 160$ yr BP (after correction for the modern air/sea $\Delta^{14}\text{C}$ difference, equivalent to 440 years²²). A sample of mixed benthic foraminifera from 25.2–25.3 m, just below reflector IIa, yielded an age of $>49,000$ yr, indicating that the fauna were derived

from pre-existing marine sediments and further confirming the glacial origin of sediments beneath the reflector. Sedimentation rates during the deglaciation averaged 4.5 m kyr^{-1} , indicating by extrapolation that marine invasion had begun and that the ice sheet had retreated from the central part of the channel by 15 kyr BP.

An oxygen isotope record from the Troll 3.1 samples (Table 1; Fig. 3b), based on analysis of the benthic foraminifera *Cassidulina laevigata* after 10.5 kyr BP and *Nonion labradoricum* before, is currently the best dated and highest resolution record spanning the last deglaciation in the Norwegian Sea. The oxygen isotope ratios of foraminifera reflect the isotopic composition of the water in which the foraminifera lived, influenced by global ice volume and local salinity changes, and the temperature at which the foraminifera precipitated their tests. Isotope values in Troll 3.1 between 14.7 and 13.5 kyr BP are as low as early Holocene values (10–9 kyr), but a 0.65‰ decrease in $\delta^{18}\text{O}$ is expected between 14 and 9.5 kyr because of the reduction in global ice volume²³. Because the temperature of the Norwegian Sea²⁴ and global ice volume²³ had changed little since the LGM, the low isotope values between 14.7 and 13.5 kyr BP must record low-salinity conditions within the channel due to ice-sheet discharge (meltwater and melting icebergs). Abrupt shifts to lower $\delta^{18}\text{O}$ at 12.7–12.4 kyr and 10–9.7 kyr BP are associated with evidence of a warming at about those times²⁴, and may have been amplified by the local salinity effects of increased ice-sheet melting rates. Periods of increasing $\delta^{18}\text{O}$ that occurred 13.4–12.8 kyr and 12.0–11.7 kyr BP are associated with increased abundance of cooler fauna in Troll 3.1 (S.J.L., manuscript in preparation) and possibly with reduced local ice-sheet melting rates. These cooling events may correspond with the Oldest and Older Dryas (respectively) climate oscillations previously documented on adjacent land masses^{13,25–28}. Lower $\delta^{18}\text{O}$ values at 11–10.5 kyr are associated with a cold water, low-salinity fauna (S.J.L., manuscript in preparation), and with terrestrial evidence for cooling and ice-sheet regeneration during the first half of the Younger Dryas²⁵. This leads us to conclude that salinities were reduced in the channel at this time, possibly as a result of meltwater discharge from a Baltic ice lake²⁷. Attainment of the lowest $\delta^{18}\text{O}$ values at about 8 kyr BP corresponds with the opening of the English Channel and increased flow of warm Atlantic water into the North Sea²⁹.

On the basis of an AMS-dated planktonic $\delta^{18}\text{O}$ record from core PS21295-4 in the Fram Strait (left curve in Fig. 3c), it has been suggested that a marine-based ice sheet situated over the Barents Shelf retreated catastrophically between 15 and 14.5 kyr BP (ref. 6). Two of us have since AMS-dated two additional planktonic $\delta^{18}\text{O}$ records in cores from the Bear Island Slope (V27-60) and Denmark Strait (V28-14)³⁰ showing that $\delta^{18}\text{O}$ ratios and therefore that salinities were reduced throughout the Norwegian Sea at this time (Fig. 3c). Low $\delta^{18}\text{O}$ ratios in sediments deposited just above till in Troll 3.1 provide the first direct confirmation that ice-sheet discharge caused the widespread

isotope anomaly. Previous glacial geological studies called for a rapid, ice-sheet retreat of 60 km from the continental-shelf edge in an area close to the proposed confluence of the Barents Shelf and Fennoscandian ice sheets sometime between $16,150 \pm 330$ and $13,700 \pm 400 \text{ yr BP}$ ³¹. The fact that planktonic $\delta^{18}\text{O}$ measured in three widespread deep Norwegian Sea cores was the same before 15 kyr BP and then quickly and simultaneously shifted to lower values (Fig. 3c), strongly suggests that retreat in this area was synchronous with retreat in the Norwegian Channel and thus that the onset of shelf deglaciation began simultaneously between 60° and $\sim 75^\circ \text{ N}$.

Our data have several important climatological implications. First, initial (apparently widespread) deglaciation began $\sim 2 \text{ kyr}$ too early to be readily explained by the heating effects of increased poleward advection of warm Atlantic surface waters²³. Jones and Keigwin¹ suggested that the onset of ice sheet retreat resulted from an autocyclic, glacio-isostatic lowering of the ice-sheet bed but recent reconstructions suggest that the ice-sheet margins were relatively thin^{34,35}, making such a response unlikely. Furthermore, additional dating of the Barbados sea-level record makes the appeal to an autocyclic mechanism unnecessary as up to 8 m of eustatic sea-level rise apparently occurred before 15–14.5 kyr BP (ref. 32). In addition, according to a recent recalibration of the ^{14}C timescale⁷, the 15,000 ^{14}C yr BP deglaciation event now corresponds to a time of 18,000 calendar yr BP, when summer insolation at 60° N began to rise rapidly above LGM levels³³. The low-elevation, low-gradient ice-sheet profiles now being proposed would be intrinsically more sensitive to increasing insolation than the steep-gradient (East Antarctic type) ice-sheet profiles previously reconstructed for the area⁴², because a given insolation-induced rise of the snowline would throw proportionally larger areas of the ice-sheet margin into the ablation zone. Thinning due to surface ablation would destabilize the ice-sheet margin by reducing its tendency to remain grounded on the shelf and buoyant forces would also increase due to eustatic sea level rise. Once buoyant decoupling of the ice sheet from the bed is achieved, iceberg calving will lead to rapid, irreversible retreat until the grounding line again shoals³⁶. Our isotope data suggest that discharge through the channel continued until 13.5 kyr BP (Fig. 3b), even though the deepest part of the channel had been rapidly deglaciated 15 kyr BP. This is compatible with glaciological modelling³⁷ and field evidence²⁸ that an ice stream developed near the mouth of the Baltic following retreat of the ice margin from the channel. Current chronologies suggest that the ice stream persisted until $\sim 13.5 \text{ kyr BP}$ ²⁸, in complete agreement with the isotope data.

A final implication of our data is that, according to the record from core V28-14 in the Denmark Strait (Fig. 3c), meltwater and icebergs were being routed from the Norwegian Sea into the northern North Atlantic beginning $\sim 15 \text{ kyr BP}$, a time when earlier studies indicate that the subpolar Atlantic was micropalaeontologically 'barren'³⁸ and the subtropical Atlantic was briefly much colder than during the LGM⁸. Cooling at that time

FIG. 2 Generalized late Quaternary cross-section of the Norwegian Channel (after ref. 18), showing the position of Troll 3.1. Symbols refer to (1), acoustic blanking; (2), silts, often with layers of sand and clay; (3) soft, laminated clays; (4), discontinuously overconsolidated, massive clay-rich gravelly diamict; (5), continuously overconsolidated massive clay-rich, stoney diamict; (6), crystalline bedrock; (7), sand.

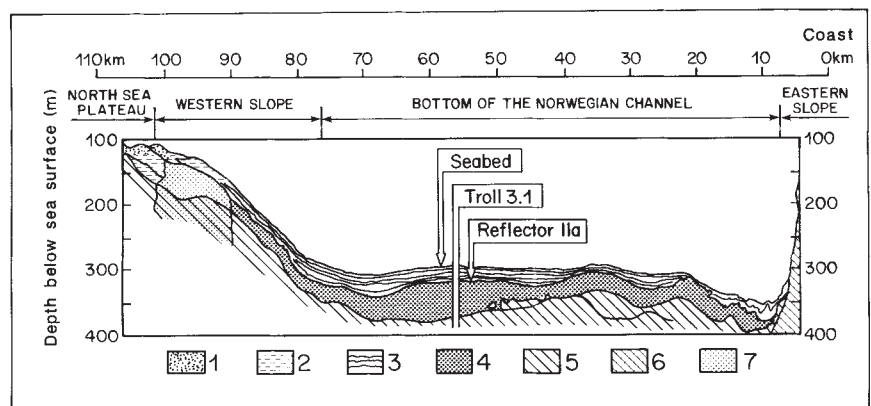
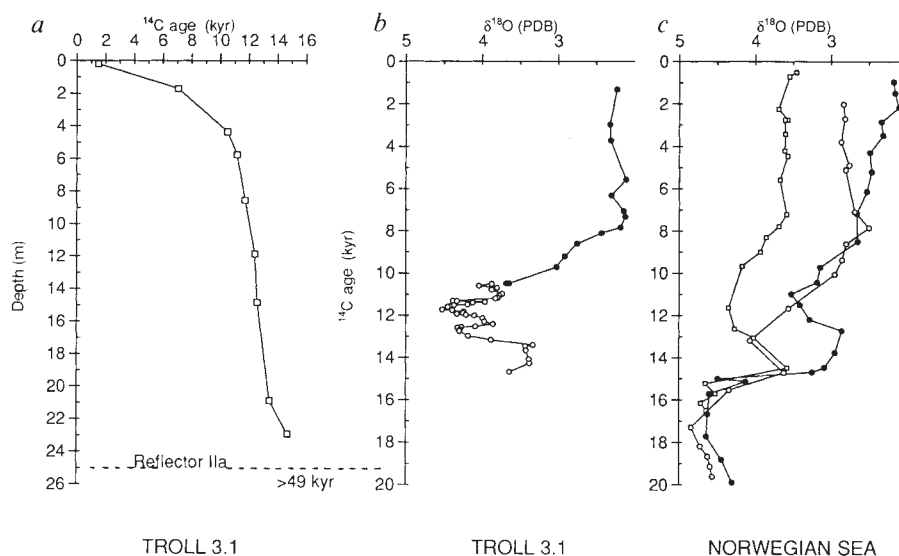


FIG. 3 a, AMS ^{14}C results against depth for Troll 3.1. All ages have been reservoir corrected by 440 yr²⁵. According to ref. 5 there is currently no evidence to suggest a change in this correction during the deglaciation. $\delta^{18}\text{O}$ in *C. laevigata* (●) and *N. labradoricum* (○) plotted against age for Troll 3.1 (b) using the age–depth relationship in (a). c, $\delta^{18}\text{O}$ in *N. pachyderma sin.* for Fram Strait core PS2195-4 (left curve), core V27-60 from the base of the Bear Island Fan, and Denmark Strait core V28-14 (right curve). $\delta^{18}\text{O}$ results and AMS ^{14}C age control for PS21295-4 are from ref. 6. AMS ^{14}C age control for cores V27-60 and V28-14 are from ref. 30, and $\delta^{18}\text{O}$ data are from refs 40 and 41, respectively.



is also registered by a brief interval of high planktonic $\delta^{18}\text{O}$ in AMS-dated cores from the Bermuda Rise⁵, and is associated with Cd/Ca vectors in benthic foraminifera reaching typical Pacific values, thus indicating reduced or eliminated production of North Atlantic Deep Water¹⁰. The spatial pattern of these variables at 15 kyr is similar to that observed during the Younger Dryas^{9,10} and suggests a similar climate response at the two times, and possibly similar climate forcing. The climatic cooling wrought by ice-sheet discharge at 15 kyr BP now provides a satisfying explanation for why most of the southern margin of the Laurentide ice sheet re-advanced at that time^{1,2,3,9}, whereas the higher-latitude Fennoscandian ice sheet was in retreat. □

40. Labeyrie, L. & Duplessy, J.-C. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **50**, 217–240 (1985).
41. Kellogg, T. in *Climatic Changes on a Yearly to Millennial Basis* (eds Morner, N.-A. & Karlen, W.) 123–133 (Reidel, Dordrecht, 1984).
42. Vorren, T. O. *Boreas* **6**, 247–257 (1977).

ACKNOWLEDGEMENTS. We thank W. F. Ruddiman and R. G. Fairbanks for stimulating discussions.

Metamorphism of eucrite meteorites studied quantitatively using induced thermoluminescence

J. David Batchelor & Derek W. G. Sears

Cosmochemistry Group, Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, USA

EUCRITE meteorites¹ are especially important in studies of the early Solar System because they are the simplest and most ancient products of a process that was widespread in the inner Solar System^{2,3}: basaltic volcanism. They are also the meteorites for which there is least doubt of an asteroidal origin^{4,5}. After volcanism the eucrites experienced a period of metamorphism⁶, either inside the asteroid as it cooled from igneous temperatures⁷ or on the surface of the asteroid as a result of impact heating⁸. Induced thermoluminescence studies provide a new and quantitative means of determining relative metamorphic intensities for these meteorites. Using this technique, we show that the eucrites constitute a continuous metamorphic series and not, as commonly assumed, two groups of metamorphosed and non-metamorphosed meteorites⁹. These studies are the first application of the induced thermoluminescence technique to igneous rocks and we suggest that the method may well have application to other basalts.

We suspected that induced thermoluminescence (TL) might be useful in investigating metamorphism of eucrites because a variety of studies have shown TL to be of great value in understanding metamorphism of chondritic meteorites^{10,11}. Chondrites, however, are essentially 'cosmic sediments', in which the feldspar—the mineral responsible for the TL—is almost entirely metamorphic in origin¹². In eucrites, the feldspar is igneous in origin, but many of the meteorites are surface breccias and impact reworking may have converted some of the feldspathic material to glass¹³. Metamorphism, depending on its timing relative to brecciation and the relative intensities of the two

Received 23 August; accepted 6 December 1990.

1. Dreimanis, A. *Geogr. Phys. Quat.* **31**, 37–51 (1977).
2. Clayton, L. & Moran, S. *Quat. Sci. Rev.* **1**, 55–82 (1982).
3. Dyke, A. & Prest, V. *Geogr. Phys. Quat.* **61**, 237–263 (1987).
4. Broecker, W. et al. *Nature* **341**, 318–321 (1989).
5. Keigwin, L. & Jones, G. *Deep Sea Res.* **36**, 846–867 (1989).
6. Jones, G. & Keigwin, L. *Nature* **226**, 56–59 (1988).
7. Bard, E. et al. *Nature* **345**, 405–410 (1990).
8. Bard, E. et al. *Nature* **328**, 791–794 (1987).
9. Boyle, E. A. & Keigwin, L. D. *Nature* **330**, 35–40 (1987).
10. Keigwin, L. D. et al. *J. geophys. Res.* (in the press).
11. Sejrup, H.-P. et al. *Norsk Geol. Tidsskr.* **69**, 111–124 (1989).
12. Petersen, K.-S. in *Climatic Changes on a Yearly to Millennial Basis* (eds Morner, N.-A. & Karlen, W.) 63–68 (Reidel, Dordrecht, 1984).
13. Paus, A. *Boreas* **17**, 113–139 (1988).
14. Paus, A. *J. Quat. Sci.* **4**, 223–242 (1989).
15. Sejrup, H.-P. et al. *J. Quat. Sci.* **24**, 35–58 (1987).
16. Cameron, T. D. et al. *J. geol. Soc. London* **144**, 43–58 (1987).
17. Andersen, B. in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T.) 3–54 (Wiley, New York 1981).
18. Butenko, G. & Østmo, S.-R. *Geol. Soc. spec. Pub.* (in the press).
19. Rise, L. & Roekongen, K. *Mar. Geol.* **58**, 287–315 (1984).
20. Van Weering, T. C. E. *Geol. Min.* **62**, 319–328 (1983).
21. Andersen, E.-S. thesis, Univ. Oslo (1990).
22. Mangerud, J. & Gulliksen, S. *Quat. Res.* **5**, 263–273 (1975).
23. Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
24. Jansen, E. & Bjørklund, K. *Boreas* **14**, 243–258 (1985).
25. Mangerud, J. et al. *Boreas* **3**, 110–127 (1974).
26. Atkinson, T. C. et al. *Nature* **325**, 587–592 (1987).
27. Bjørck, S. & Møller, P. *Quat. Res.* **28**, 1–37 (1987).
28. Ringberg, B. *Boreas* **17**, 243–263 (1988).
29. Jelgersma, S. in *The Quaternary History of the North Sea* (eds Oele, E. Schuttenheim, R. T. E. & Wiggers, A. J.) 233–248 (Acta Univ. Uppsala, Uppsala, 1979).
30. Jones, G. A. & Lehman, S. J. *Paleoceanography* (submitted).
31. Vorren, T. et al. *Boreas* **17**, 41–77 (1988).
32. Fairbanks, R. G. *Paleoceanography* (in the press).
33. Berger, A. *Quat. Res.* **9**, 139–167 (1978).
34. Nesje, A. & Sejrup, H.-P. *Boreas* **17**, 371–384 (1988).
35. Grosswald, M. *Data on Glaciol. Studies* **63**, 3–24 (1988).
36. Hughes, T. in *The Geology of North America Vol. K-3* (eds Ruddiman, W. F. & Wright, H. E.) 183–220 (Geol. Soc. Am. Boulder, 1987).
37. Boulton, G. S. et al. *J. geol. Soc. Lond.* **142**, 447–474 (1985).
38. Ruddiman, W. F. & McIntyre, A. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **35**, 145–214 (1981).
39. Broecker, W. S. & Denton, G. H. *Geochim. cosmochim. Acta* **53**, 2465–2501 (1989).

processes, might well have caused recrystallization of the feldspar¹⁴. If this is the case, metamorphism can be monitored rather precisely using induced TL of the feldspar.

Interior chips of mass ~160 mg were taken from 18 eucrites, duplicate chips being taken where possible. These were homogenized by hand-crushing in an agate mortar, and three 4-mg aliquots were placed in shallow copper pans and drained of their natural TL by brief heating to 500 °C. The TL induced by a 15-TBq ⁹⁰Sr β -particle source was then measured by heating at 7.5 °C s⁻¹ in TL apparatus equipped with blue and infrared bandpass filters (Corning 7-59 and 4-69; ref. 15). Data for three aliquots were averaged. The H3.8 ordinary chondrite Dhajala was used as a normalization standard and instrument check.

The intensity of the TL at maximum light production (referred to as TL sensitivity) is plotted in Fig. 1 for each eucrite, ranked in descending order on a logarithmic scale. Duplicate data for different chips from a single eucrite are connected by tie-lines. Reid and Barnard divided the eucrites into 'equilibrated' and 'unequilibrated' groups, suggesting that the former were metamorphosed whereas the latter were not⁹. More recently, Takeda *et al.*¹⁶ sorted the eucrites and several clasts separated from eucrites into six 'types' according to metamorphism, but their scheme has not been widely adopted. Both these authors, and several others, have used chemical zoning and exsolution in the pyroxenes to assess metamorphism in these meteorites. In the little-metamorphosed eucrites, igneous zoning is present, but for the more metamorphosed eucrites it is absent and exsolution of augite has occurred. The criteria used by Takeda *et al.* for their six types are listed in Table 1. We find that the TL sensitivities of the eucrites are spread uniformly over a 17-fold range. There is no indication that the eucrites formed in two discrete environments, corresponding to the equilibrated and unequilibrated categories of Reid and Barnard, but rather, as suggested by Takeda *et al.*, that they formed in a single environment that permitted a range of metamorphic conditions.

If we subdivide our TL range into intervals based on five divisions per decade on a log scale (as indicated in Fig. 1), we find agreement with Takeda *et al.*'s types in four out of the five cases for which comparison is possible. (Unfortunately, most of Takeda *et al.*'s samples are not readily available in the United States.) We can therefore propose a scheme based on TL sensitiv-

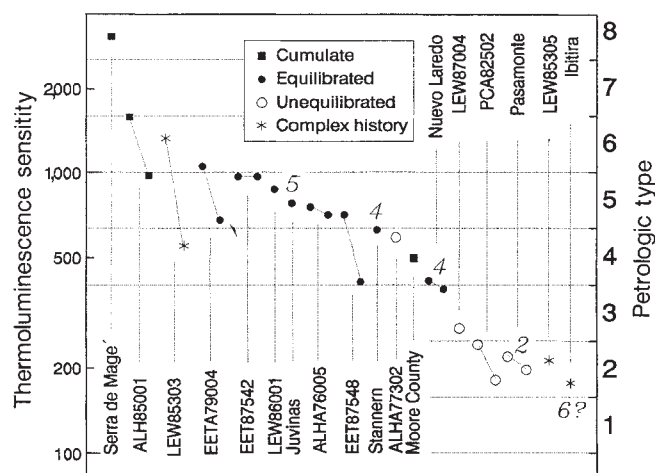


FIG. 1 Thermoluminescence (TL) sensitivity and petrologic type of 18 eucrites. The H3.8 ordinary chondrite Dhajala corresponds to a sensitivity of 1,000. Duplicate chips are connected by tie-lines. Reid and Barnard's⁹ 'equilibrated' and 'unequilibrated' categories are indicated by symbols^{2,18,30,31}, and the types assigned by Takeda *et al.*¹⁶ on the basis of pyroxene characteristics are indicated by the numbers over the symbols. There is agreement between the present classification and those assigned by Takeda *et al.*¹⁶, and the division between the unequilibrated and equilibrated categories lies between petrologic types 3 and 4. TL sensitivity is determined primarily by the metamorphism experienced by the meteorites; the data indicate that the eucrites form a continuous metamorphic series. The eucrites LEW85303, LEW85305 and Ibitira have had unusual histories, involving intense shock, which has almost certainly affected their TL data.

ity and pyroxene characteristics for metamorphic subdivision, in which eight types are required to cover the observed range (Table 1). For samples for which duplicates were available, type assignments are based on mean TL sensitivity values. Type assignments for duplicates never differ by more than two, and typically fall within the same type or differ by only one. Some of these differences may be caused by mineralogical

TABLE 1 Metamorphic classification of eucritic meteorites

Petrologic type	Thermoluminescence sensitivity (Dhajala = 1,000)	Pyroxene characteristics*	Meteorites
1	<160	Extensive Mg-Fe zoning trend in pyroxene and Fe-rich augite.	Y-75011†, clasts in Y-74450† and Y75015†
2	160–250	Fe-rich pyroxene not seen, Mg-Fe-Ca trend preserved.	Pasamonte‡, Ibitira, PCA82502, LEW85305, clasts in Y-74159†
3	250–400	Mg-Fe zoning mostly disappeared, Fe-Ca trend from core to rim preserved.	LEW87004, Y-790226†
4	400–630	Trend from host to exsolved augite dominant, original Fe-Ca trend can be recognized.	Stannern‡, Nuevo Laredo‡, Millbillillie‡, Moore Co., EET87548, ALHA77302
5	630–1,000	All zoning has disappeared, exsolution trend dominant.	Juvinas‡, Haraiya†, Sioux Co.†, LEW85303, EETA79004, EET87542, LEW86001, ALHA76005
6	1,000–1,600	Partial to complete inversion of pigeonite to orthopyroxene.	ALH85001
7	1,600–2,500	Partial to complete inversion of pigeonite to orthopyroxene.	—
8	>2,500	Partial to complete inversion of pigeonite to orthopyroxene.	Serra de Magé

* Descriptions of types 1–5 from Takeda *et al.*¹⁶.

† Assignments from Takeda *et al.*¹⁶.

‡ Assignments based on the present TL data and Takeda *et al.*'s petrographic descriptions.

Other assignments based on TL sensitivity.

Ibitira is an unusual eucrite which Takeda *et al.* "tentatively" identified as type 6.

heterogeneity in these coarse-grained rocks, but especially interesting is the possibility of differences in metamorphism experienced by different clasts within a meteorite. This would suggest that brecciation post-dated metamorphism, and was found to be the case for the related howardite meteorites¹⁷. The present data, however, indicate that, despite these small variations, petrologic types may be assigned to the whole-rock by both petrographic and TL methods with a high degree of confidence.

The specimen whose assignment differs according to the schemes of Takeda *et al.* and the present work is Ibitira, which the previous authors "tentatively" assigned to type 6. This, and presumably LEW85305 to which it is petrographically similar¹⁸, have experienced unusual histories involving intense shock and subsequent metamorphism¹⁹.

Two of the three cumulate eucrites in the present study have the highest TL sensitivities. The thermal history of cumulate eucrites is especially interesting because they seem to have relatively young radiometric ages²⁰⁻²⁶, and there is considerable debate as to whether this is because they formed in separate magma bodies or because they cooled especially slowly^{20,23}. Our data support the latter conclusion because the high TL sensitivity of two of the cumulate eucrites suggests particularly prolonged metamorphism.

We suspect that the increase in TL sensitivity in response to parent-body metamorphism is due to the increase in the degree of crystallinity of the feldspar. Cathodoluminescence petrography shows that the feldspars in the Juvinas type 5 eucrite are brighter and more uniform in their luminescence than those in the Pasamonte type 2 eucrite. Processes on the surface of the parent body probably cause some of the feldspar to be converted to glass, and metamorphism reverses this process. We cannot, however, rule out other mechanisms involving trace elements or more subtle structural changes. In the case of the related howardite class, we have observed changes in TL sensitivity that correlate with bulk aluminium content³². The howardites, however, are breccias of Al-poor diogenite and Al-rich eucrite material, whereas the eucrites in the present study are, to a good approximation, breccias of one type of material.

Additional information on the metamorphic conditions experienced by the eucrites is provided by the temperature at which the TL light production reaches a maximum. Figure 2 shows the TL 'glow curves' (light production as a function of temperature as the sample is heated) for two representative eucrites (ALHA76005 and LEW87004) and a very unusual eucrite (LEW85303). The glow curves consist of two peaks, at $113 \pm 11^\circ\text{C}$ and $192 \pm 15^\circ\text{C}$. These two peaks are thought to be indirectly related to ordered and disordered feldspar respectively¹⁰. LEW85303 was the only meteorite in the present study that showed no sign of a 113°C peak. Annealing of four eucrites (a cumulate, an equilibrated eucrite, an unequilibrated eucrite and LEW85303) in an inert atmosphere for 100 h showed that at or just below 800°C , the 113°C peak moves to 160 – 200°C , consistent with the idea that ordering of the feldspar crystal lattice affects peak temperature. As all eucrites in this study except LEW85303 had a dominant peak at 113°C , or had two peaks of comparable heights, we conclude that most eucrites cooled slowly below 800°C , but that LEW85303 has been heated to $>1,000^\circ\text{C}$. Ar-Ar data for LEW85303 also show evidence of a major degassing event at about 1 Gyr (ref. 27), and there is petrographic evidence that this meteorite has experienced a shock event^{28,29}.

Our results indicate that induced thermoluminescence studies may have much wider application than previously realized. The technique has been of considerable value in understanding metamorphism in chondrites, but these are very unusual materials in which the feldspar has a novel origin and history. The eucrites, on the other hand, are igneous rocks not unlike terrestrial and lunar basalts. The successful application of induced TL to studies of the thermal histories of the eucrites suggests that the technique may be applied usefully to a wide variety of basalt types for which dry, low-level metamorphism has occurred, and preliminary work on terrestrial³² and lunar³³ samples is encouraging. □

Received 27 July; accepted 3 December 1990.

- Duke, M. B. & Silver, L. T. *Geochim. cosmochim. Acta* **31**, 1637–1665 (1967).
- Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).
- McSweeney, H. Y. A. *Rev. Earth planet. Sci.* **17**, 117–140 (1989).
- Chapman, C. R. *Geochim. cosmochim. Acta* **40**, 701–719 (1976).
- Consolmagno, G. Y. & Drake, M. J. *Geochim. cosmochim. Acta* **41**, 1271–1282 (1977).
- Miyamoto, M., Duke, M. B. & McKay, D. S. *Proc. 15th Lunar planet. Sci. Conf.* C629–C635 (1985).
- Miyamoto, M. *Lunar planet. Sci.* **14**, 505–507 (1983).
- Delaney, J. S., Prinz, M., Nehru, C. E. & O'Neill, C. *Proc. 13th Lunar planet. Sci. Conf.* A339–A352 (1982).
- Reid, A. M. & Barnard, B. M. *Lunar planet. Sci.* **10**, 1019–1022 (1989).
- Sears, D. W., Grossman, J. N., Melcher, C. L., Ross, L. M. & Mills, A. A. *Nature* **287**, 791–795 (1980).
- Guimon, R. K., Weeks, K. S., Keck, B. D. & Sears, D. W. G. *Nature* **311**, 363–365 (1980).
- Van Schmus, W. R. & Wood, J. A. *Geochim. cosmochim. Acta* **31**, 747–765 (1967).
- Labotka, T. C. & Papike, J. J. *Proc. 11th Lunar planet. Sci. Conf.* 1103–1130 (1980).
- Hag, M., Hasan, F. A., Sears, D. W. G., Moore, C. B. & Lewis, C. F. *Geochim. cosmochim. Acta* **53**, 1435–1440 (1989).
- Weast, R. C. (ed.) *CRC Handbook of Chemistry and Physics* 66th Edn (CRC Press, Boca Raton, 1986).
- Takeda, H. *et al. Proc. 8th Symp. Antarctic Meteor.* 181–205 (1983).
- Batchelor, J. D. & Sears, D. W. G. *Meteoritics* (in the press).
- Mason, B. in *Antarctic Meteorite News* (Johnson Space Center, Houston, Texas, 1978–1989) (ALHA76005, 2(1), 4(1); ALHA77302, 1(2), 1(3), 4(1); ALHA85001, 9(3); EETA79004, 3(3), 4(1); EET87452, 12(1); EET87548, 12(1); LEW85303, 9(3); LEW85305, 9(3); LEW86001, 10(2); LEW87004, 11(2); PCA82502, 6(2), 7(1)).
- Steele, I. M. & Smith, J. V. *Earth planet. Sci. Lett.* **33**, 67–78 (1976).
- Tera, F., Carlson, R. W. & Bockor, N. Z. *Meteoritics* **22**, 513–514 (1987).
- Jacobson, S. B. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **67**, 137–150 (1984).
- Lugmair, G. W., Scheinin, N. B. & Carlson, R. W. *Meteoritics* **10**, 300–301 (1977).
- Tera, F., Carlson, R. W. & Bockor, N. Z. *Lunar planet. Sci.* **20**, 1111–1112 (1989).
- Birk, L. L. & Allegre, C. J. *Earth planet. Sci. Lett.* **39**, 37–51 (1978).
- Carlson, R. W., Tera, F. & Bockor, N. Z. *Lunar planet. Sci.* **19**, 166–167 (1988).
- Warren, P. H., Haack, H. & Rasmussen, K. L. *Lunar planet. Sci.* **20**, 1179–1180 (1989).
- Bogard, D. D. & Garrison, D. H. *Lunar planet. Sci.* **20**, 90–91 (1989).
- Delaney, J. S. *Meteoritics* **22**, 365–366 (1987).
- Hewins, R. H. *Lunar planet. Sci.* **21**, 509–510 (1990).
- Harlow, G. E. & Klimentidis, R. *Proc. 11th Lunar planet. Sci. Conf.* 1131–1143 (1980).
- Mason, B. *et al. Smithson. Contrib. Earth Sci.* **23**, 36–38 (1979); **24**, 38–43 (1982); **26**, 41–44 (1984); **28**, 50–53 (1989).
- Batchelor, J. D. & Sears, D. W. G. *Lunar planet. Sci.* **20**, 52–53 (1989).
- Sears, D. W. G., Sears, H. & Myers, B. M. *Lunar planet. Sci.* **21**, 1121–1122 (1990).

ACKNOWLEDGEMENTS. We thank the Meteorite Working Group of NASA/NSF, R. Clarke, P. Pellas and E. Scott for samples, P. Benoit and R. Hewins for helpful discussions, H. McSweeney and A. Reid for constructive reviews and R. Stucker for technical assistance. This work is supported by NASA.

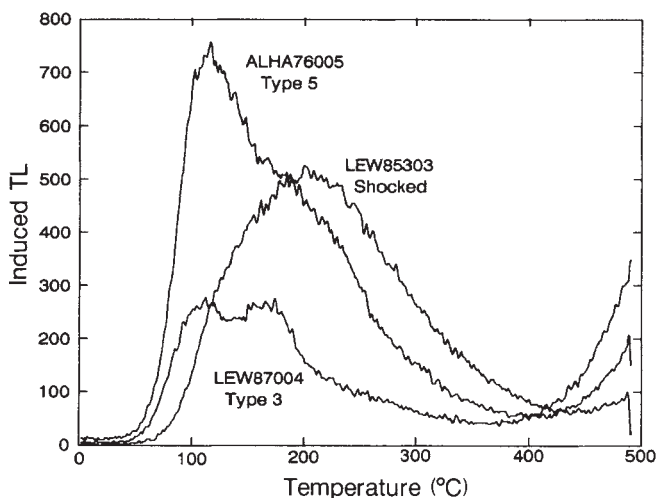


FIG. 2 Induced thermoluminescence glow curves for an equilibrated eucrite (ALHA76005, type 5), an unequilibrated eucrite (LEW87004, type 3), and the unusual shocked eucrite LEW85303. The Dhajala chondrite corresponds to 1,000. Equilibrated eucrites have a strong peak at about 110°C and a weaker peak at about 190°C , and unequilibrated eucrites have a stronger peak at 190°C . The unusual eucrite LEW85303 has a broad peak at around 200°C . A TL peak at about 110°C reflects metamorphism below 800°C , and the single peak at 200°C reflects heating to temperatures $>1,000^\circ\text{C}$.

The advantages and evolution of a morphological novelty

Craig W. Benkman & Anna K. Lindholm

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 2A9, Canada

THE role of selective agents in the origin of evolutionary novelties has been controversial¹⁻³ and has remained outside the realm of experiments. Here we experimentally determine both the benefits of a single trait and the advantages accrued during the presumed sequence of evolutionary steps leading to the fully specialized structure. By comparison of red crossbills (*Loxia curvirostra*, L.), in which the mandibular crossing has been removed, with controls and with the related but less specialized pine siskin (*Carduelis pinus* Wilson), we show the advantage of the mandibular crossing in the extraction of seeds from partially closed conifer cones. We use the natural regrowth of the mandibles to mimic the evolution of mandibular crossing from an unspecialized ancestor, and use the relationship of foraging efficiency to mandibular regrowth to determine a scheme for its (gradual) evolution.

Crossed mandibles characterize crossbills (*Loxia*) (Fig. 1) and have been pivotal in their radiation into about 25 species and subspecies^{4,5}. The pointed and decurved upper mandibles are inserted between overlapping conifer cone scales to create gaps which are then widened by lateral abduction of the lower mandibles. Once exposed, the seed is lifted out with the tongue and husked^{6,7}. An earlier experiment⁸ showed that mandibular crossing was essential for extracting seeds from closed cones. Here we further quantify this relationship for closed and open cones. We also determine the relationship between foraging efficiency and mandibular (ramphothecae) regrowth to mimic the possible steps in the evolution of mandibular crossing. The regrowth is similar to natural bill growth of young crossbills after fledging (C.W.B., unpublished observations).

Measures of performance provide an index to the degree of adaptation of features⁹⁻¹¹ such as crossed mandibles. The rate of seed (food) intake is an excellent measure of performance¹². Bill structure influences rates of seed intake by crossbills⁷ which, in turn, has been shown to affect crossbill movements, diet selection¹³, patch use¹⁴, timing of reproduction¹⁵ and presumable fitness. We measured the seed intake rates of seven captive red crossbills (*L. c. minor*) and five captive pine siskins (Fig. 1) foraging on seeds in the cones of western hemlock (*Tsuga heterophylla* [Raf.] Sarg.). This subspecies of crossbill, mostly confined to the coastal forests from Alaska to California¹⁶, has

TABLE 1 Mean ($\pm$ s.e.m.) seed extraction times

	Seed extraction times (s)†		t‡,
	Experimental	Control	
Pre-alteration			
Open cones	1.49 $\pm$ 0.15	1.13 $\pm$ 0.13	1.70
Reclosed cones	1.34 $\pm$ 0.08	1.25 $\pm$ 0.11	0.70
Closed cones	1.61 $\pm$ 0.32	1.84 $\pm$ 0.10	0.62
Post-alteration			
Open cones			
Day 1	1.92 $\pm$ 0.30	1.20 $\pm$ 0.14	1.93
6	1.66 $\pm$ 0.29	1.43 $\pm$ 0.10	0.66
Reclosed cones			
Day 1			
3	5.28 $\pm$ 0.11¶	1.33 $\pm$ 0.07	30.33***
6	5.57 $\pm$ 1.12	1.61 $\pm$ 0.32	2.92*
12	3.51 $\pm$ 0.26	1.59 $\pm$ 0.13*	5.85*
17	2.42 $\pm$ 0.21	1.59 $\pm$ 0.13*	3.13*
22	2.72 $\pm$ 0.16	1.78 $\pm$ 0.21	3.65*
29	1.96 $\pm$ 0.07	1.54 $\pm$ 0.10	2.86*
36	1.68 $\pm$ 0.21	1.24 $\pm$ 0.05	1.73
Closed cones			
Day 1			
3		3.06 $\pm$ 0.12	
6			
8		2.88 $\pm$ 0.19	
12			
17	5.21 $\pm$ 0.29¶	2.89 $\pm$ 0.09	7.73**
22	4.64 $\pm$ 0.35¶	3.30 $\pm$ 0.37	2.66*
29††	4.03 $\pm$ 0.36	2.81 $\pm$ 0.04	2.86*

Data are for four experimental and three control red crossbills foraging on western hemlock cones.

† Seed extraction time is mean total time per seed minus mean seed husking time (see refs 7, 8). After the first seed removed from the cone was swallowed, we (A.K.L.) timed the interval of the extraction and consumption of the following five seeds (10 seeds before bill alteration) to the nearest 0.1 s. Data are based on the means of 9.6 $\pm$ 0.1 cones per bird and represent a total of 991 cones. Mean seed husking time (the interval required to remove the seed wing and seed coat and swallow the kernel) was determined each day for each individual on the basis of 16.4 $\pm$ 0.7 seeds per sample, for a total of 1,717 seeds.

‡ Two-tailed t-tests: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

§ Days after final bill trimming. We were often unable to obtain foraging data on both experimental and control groups on one day. Thus controls were measured as close in time as possible to bill-altered crossbills (0–3-day difference). Days refer to day of foraging trial of the experimental group, except for closed cones on days 3 and 8.

|| Unable to remove five seeds from five cones.

¶ Based on data from three birds.

* Represent data gathered on day 15.

†† Closed cones were depleted before seed extraction time for bill-altered crossbills converged to those of controls.

FIG. 1 Lateral profiles of an unaltered red crossbill (left), a bill-altered crossbill (centre) and a pine siskin. Experimental crossbills were chosen by ranking them according to bill depth and selecting alternate ranks. The mean bill depths of three control and four experimental crossbills were 8.14 mm (range 7.94–8.43 mm) and 8.20 mm (range 7.87–8.64 mm), respectively. Siskins had a mean bill depth of 5.88 mm (range 5.73–6.06 mm). The upper and lower mandibles of the four experimental crossbills were reduced on average to 79.1 and 83.9% of their mean original lengths of 12.27 mm (s.e.m. = 0.22) and 10.08 mm (s.e.m. = 0.13), respectively. We used nail-clippers to remove most of the mandible crossing from four crossbills on one day and the following day further trimmed the mandibles. The three controls were handled similarly for the same length of time (~8 min over 2 days) without bill clipping. The bill-altered crossbills showed no more signs of stress than the controls. We applied pitch (from *Abies* species) to the bill of all crossbills to promote bill-wiping against grit-covered perches to round the cut edges of the mandibles. After bill clipping, we (C.W.B.) measured the upper and lower mandible lengths (0.01 mm) seven times with digital calipers as the mandibles regrew. A quadratic equation provided a good fit to mandible regrowth of bill-altered crossbills (upper mandible length (mm) = 9.709 + 0.085 (day) – 0.001 (day)², $r^2 = 0.998$, $F_{2,5} = 1115.87$ $P < 0.0001$; lower



mandible length (mm) = 8.446 + 0.094 (day) – 0.001 (day)², $r^2 = 0.99$, $F_{2,5} = 288.40$, $P < 0.0001$), whereas the mandible lengths of controls grew slightly (upper mandible: mm per day = 0.004 $\pm$ 0.001, $r^2 = 0.84$, $F_{1,6} = 30.80$, $P < 0.005$) or not at all (lower mandible: mm per day = –0.0003 $\pm$ 0.005, $r^2 = 0.001$, $F_{1,6} = 0.005$, $P > 0.90$). Thirty-three days after bill clipping and three days before the last experiments the upper and lower mandibles were 96.9 and 101.2% of their original lengths, respectively. We did not determine repeatability of bill measurements during the study, but mean re-measurement error on 42 study siskins after one week was 0.04 mm for the upper mandible and 0.08 mm for the lower mandible. The upper mandible was measured from the anterior edge of the right nares to the tip of the mandible. The lower mandible was measured from the base where the rami meet to the tip of the mandible.

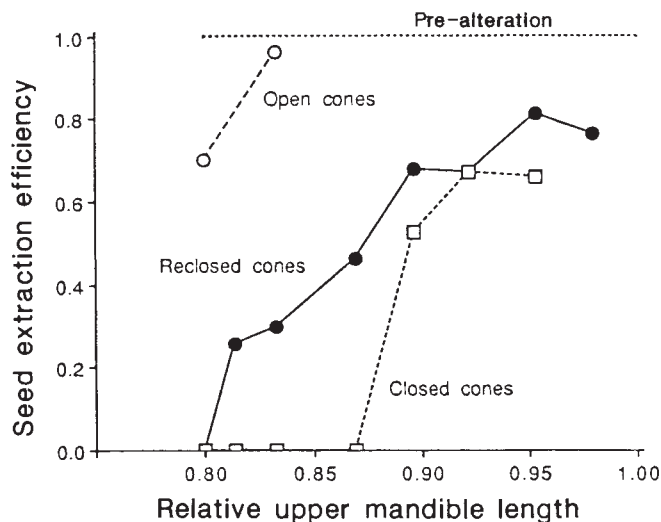


FIG. 2 Seed extraction efficiency of bill-altered crossbills relative to controls during mandible regrowth. The abscissa represents the ratio of the upper mandible length as estimated from the regression for regrowth (see Fig. 1 legend) to that measured before clipping. Measurements of the upper mandible were used because they are more accurate than those of the lower mandible (see Fig. 1 legend). The ordinate represents the reciprocal of mean seed extraction time for bill-altered crossbills divided by the same for controls, standardized to the same ratio before bill ablation. Foraging trials were conducted on birds that had fasted for ≥ 16 h. Open cones: scales had separated, thus exposing seeds (common during dry periods from autumn to spring). Reclosed cones: scales had separated, but were closed from moisture due to rain or humidity (autumn to spring). Closed cones: scales had yet to separate (summer). All cones (19.5–22.5-mm long) were gathered from two trees.

a bill that is adapted for foraging specifically on western hemlock seeds (C.W.B., in preparation). Siskins also forage on hemlock seeds but have a more generalized diet^{17,18}. We used cones representing summer foods (closed green cones) and winter foods (dry open cones and cones reclosed from moisture).

Mandibular crossing is essential for extracting seeds from between closed cone scales and increases in mandibular crossing enhance foraging efficiency on reclosed cones (Fig. 2, Table 1). The bill-altered crossbills could consistently remove seeds from closed cones, whose scales are difficult to separate, only after the upper mandibles had regrown to 90% of their original length. As the experimentally altered mandibles regrew to their original lengths, times to extract seeds from closed and reclosed cones converged on extraction times of controls (Fig. 2). Siskins could not consistently extract seeds from closed or reclosed cones during foraging trials. Thus, slightly crossed mandibles are advantageous for foraging for seeds in reclosed cones. As mandibular crossing increases, foraging efficiency increases on reclosed cones and seeds between tightly closed scales of closed cones become accessible.

Crossbills with and without mandibular crossings do not differ in their efficiency of seed extraction from open cones (Table 1). The rapid increase in efficiency on open and reclosed cones between the first and second post-alteration trials (Fig. 2) may reflect crossbills learning to use their shortened bills. Seed husking times of bill-altered crossbills do not differ from controls at any time on any cone (t -tests, $P > 0.20$). The absence of significant differences in these comparisons indicates that bill alteration did not injure the birds.

Siskins extract seeds from open cones at similar rates (1.02 ± 0.11 s per seed (mean $\pm$ s.e.m.); sample size, $N = 5$ birds) to crossbills before bill alteration (1.33 ± 0.12 s per seed, $N = 7$ birds; $t = 1.88$, d.f. 10, $P > 0.05$) and may out-compete crossbills for these seeds. On the basis of measured foraging rates, average number of seeds consumed per cone (mean = 15 and 9 seeds

per cone for crossbills and siskins, respectively), and assuming equal travel time between cones (3.4 ± 0.87 s, $N = 9$, for crossbills in field (C.W.B., unpublished data)), identical assimilation efficiencies, estimated daily energy expenditures¹⁹ and mean lean body mass (26.7 g for 7 crossbills, 12.5 g for 5 siskins), we estimate that siskins require only 74% as much time as crossbills to meet their daily energy demands when foraging exclusively on seeds in open hemlock cones. These results indicate that access to seeds in open cones has not been critical in the evolution of the mandible crossing.

We conclude that directional selection for increases in mandibular crossing enhances foraging efficiency on reclosed or incompletely opened conifer cones and is responsible for the origin of crossbills from an unspecialized ancestor. The subsequent diversification of crossbills reflects adaptations for harvesting seeds from various conifers. The ancestral crossbill may have been a pine siskin-like finch (*Carduelis*)^{20,21} which laterally abducted its lower mandible²² to improve access to seeds such as those between separated conifer cone scales⁸. Isolated occurrences of slight mandibular crossings are found occasionally in many species (C.W.B., unpublished observations). Our results indicate that such an occurrence in an unspecialized crossbill ancestor would lead to selection for progressively greater mandibular crossing to improve foraging efficiency on moisture-closed winter cones. Selection would have ensued for increased mandibular crossing to exploit a resource unavailable to less specialized finches, especially in winter which is the time when food is most limiting^{5,13}. The shape or curvature of the mandibles also affects foraging efficiency^{7,8,23}. Curvature of the mandibles could be altered presumably by differential growth within the rhamphotheca. As mandibular crossing increased, seeds in progressively more closed cones, such as those of maturing cones of summer, would be harvested. Hence, the mandibular crossing secondarily enabled specialization on seeds in conifer cones throughout the year. □

Received 9 October; accepted 13 November 1990.

1. Frazzetta, T. H. *Complex Adaptations in Evolving Populations* (Sinauer Associates, Sunderland, MA, 1975).
2. Futuyma, D. J. *Evolutionary Biology*, 2nd edn (Sinauer Associates, Sunderland, MA, 1986).
3. Arnold, S. J. et al. in *Complex Organismal Functions: Integration and Evolution in Vertebrates* 403–433 (Wiley, New York, 1989).
4. Griscom, L. *Proc. Boston Soc. nat. Hist.* **41**, 77–210 (1937).
5. Newton, I. *Finches* (Collins, 1972).
6. Tordoff, H. B. *Condor* **56**, 346–358 (1954).
7. Benkman, C. W. *Wilson Bull.* **99**, 351–368 (1987).
8. Benkman, C. W. *Ibis* **130**, 288–293 (1988).
9. Arnold, S. J. *Am. Zool.* **23**, 347–361 (1983).
10. Emerson, S. B. & Arnold, S. J. in *Complex Organismal Functions: Integration and Evolution in Vertebrates*, 295–314 (Wiley, New York, 1989).
11. Bock, W. J. & Wahler, G. v. *Evolution* **19**, 269–299 (1965).

12. Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton University Press, 1986).
13. Benkman, C. W. *Ecol. Monogr.* **57**, 251–267 (1987).
14. Benkman, C. W. *Ornis Scandinavica* **20**, 65–68 (1989).
15. Benkman, C. W. *Auk* **107**, 376–386 (1990).
16. Gabrielson, I. N. & Lincoln, F. C. *Birds of Alaska* (Stackpole, Harrisburg, PA, 1959).
17. Austin, O. L. *U. S. nat. Mus. Bull.* **237**, 424–447 (1968).
18. Martin, A. C., Zim, H. S. & Nelson, A. L. *American Wildlife and Plants* (Dover, New York, 1951).
19. Walsberg, G. E. *Avian Biology*, Vol. 7 161–220 (Academic Press, New York, 1983).
20. Raikow, R. J. *Bull. Carnegie Mus. nat. Hist.* **7**, 1–43 (1978).
21. Martin, J. A. & Johnson, N. K. *Condor* **88**, 409–420 (1986).
22. Tallman, D. A. & Zusi, R. L. *Auk* **101**, 155–158 (1984).
23. Benkman, C. W. *Evolution* **43**, 1324–1330 (1989).

ACKNOWLEDGEMENTS. We thank D. Schluter for providing laboratory space and D. Schluter and W. J. Bock for helpful comments on this manuscript. C.W.B. was supported by a NSF-NATO Postdoctoral Fellowship.

A physiological role for endogenous zinc in rat hippocampal synaptic neurotransmission

Xinmin Xie & Trevor G. Smart*

Department of Pharmacology, School of Pharmacy,
29–39 Brunswick Square, London WC1N 1AX UK

*To whom correspondence should be addressed

THE mammalian central nervous system (CNS) contains an abundance of the transition metal zinc, which is highly localized in the neuronal parenchyma^{1–4}. Zinc is actively taken up^{5,6} and stored in synaptic vesicles in nerve terminals^{7–10}, and stimulation of nerve fibre tracts that contain large amounts of zinc, such as the hippocampal mossy fibre system⁴, can induce its release^{11–13}, suggesting that it may act as a neuromodulator. The known interaction of zinc with the major excitatory and inhibitory amino-acid neurotransmitter receptors in the CNS supports this notion^{14–16}. That zinc has a role in CNS synaptic transmission, however, has so far not been shown. Here we report a physiological role for zinc in the young rat hippocampus (postnatal, P3–P14 days). Our results indicate that naturally occurring spontaneous giant depolarizing synaptic potentials (GDPs) in young CA3 pyramidal neurones, mediated by the release of GABA (γ -aminobutyric acid)¹⁷, are induced by endogenously released zinc. These synaptic potentials are inhibited by specific zinc-chelating agents. GDPs are apparently generated by an inhibitory action of zinc on both pre- and postsynaptic GABA_B receptors in the hippocampus. Our study implies that zinc modulates synaptic transmission in the

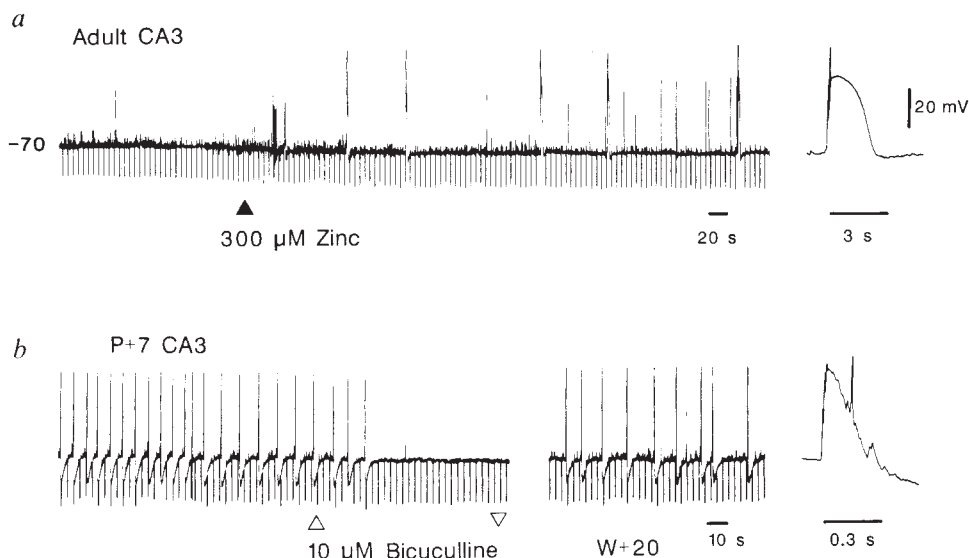
immature hippocampus, a finding that may have implications for understanding benign postnatal seizures in young children suffering with acute zinc deficiency¹⁸.

Intracellular recording from adult rat CA3 hippocampal neurons ($P > 90$) revealed mainly spontaneous inhibitory postsynaptic potentials (i.p.s.ps) and occasional action potentials (Fig. 1a). Bath-application of zinc (50–300 μ M) induced the appearance of large spontaneous depolarizations (amplitude 29 ± 6 mV, duration 2–4 s) usually followed by an after-hyperpolarization ($n = 47$ cells; Fig. 1a). The membrane conductance was increased during the depolarization, mainly owing to the activation of GABA_A receptors, as these events had identical reversal potentials to GABA responses recorded from the same cells (-53 ± 7 mV, potassium acetate, or -30 ± 3 mV, KCl filled microelectrodes) and were also blocked by the GABA_A receptor antagonist, bicuculline (10 μ M). These zinc-induced depolarizing potentials appeared similar to transient, spontaneous depolarizations which occur naturally in immature CA3 pyramidal neurons from young rats (P3–P14) not subjected to any previous pharmacological manipulation¹⁷. These regular depolarizations (GDPs) (amplitude 35 ± 9 mV, duration 200–500 ms; $n = 54$) were also inhibited by bicuculline (10 μ M), again implying mediation by GABA_A receptors (Fig. 1b). The involvement of neurotransmitter release in generating spontaneous GDPs was demonstrated by their blockade in the presence of 1 μ M tetrodotoxin, or by raising extracellular calcium (4 mM) and magnesium (6 mM) concentrations, which blocks polysynaptic transmission.

Endogenous zinc in the mossy fibre nerve terminals might be responsible for the spontaneous GDPs in the immature hippocampus. We investigated this by using selective zinc-chelating agents based on heterocyclic pyridinones¹⁹, to remove the influence of endogenous zinc on the young CA3 pyramidal cells. These compounds were selected according to their partition

FIG. 1 a, Zinc induces giant depolarizing synaptic potentials (GDPs) in adult hippocampal CA3 neurons. Chart record of an intracellular recording of membrane potential adjusted with DC current injection to -70 mV with superimposed hyperpolarizing electrotonic potentials evoked by -0.3 nA, 300 ms, 0.2 Hz current pulses. Control trace illustrates synaptic activity (i.p.s.ps) on CA3 neurons and with addition of 300 μ M zinc, the membrane potential hyperpolarizes with a small decrease in the input conductance. After 2 min GDPs with increased spontaneous action potential firing appear. The right-hand trace illustrates one spontaneous GDP on an expanded timescale. b, Young, immature CA3 neurons (P7) exhibit regularly occurring spontaneous GDPs in the absence of exogenous zinc. The GDPs and i.p.s.ps are blocked by 10 μ M bicuculline, which was reversible after a 20-min wash-out (W + 20). The right-hand trace shows a typical spontaneous GDP from the same neuron, which has a shorter duration compared to the adult GDP in a. Membrane potential (adjusted with DC current injection), -70 mV. Note the different time calibrations in a and b.

METHODS. Experiments were performed on transverse rat hippocampal slices prepared using a McIlwain tissue chopper producing 400 μ m-thick adult slices and 600 μ m-thick young postnatal slices. Adult animals were at least P90 and young rats were P3–P21. The slices were held submerged between two Nylon meshes and superfused with a Krebs' solution containing (mM): NaCl, 118; KCl, 3; CaCl₂, 2; MgCl₂, 2; NaHCO₃, 25; D-glucose, 11, bubbled with 95% O₂/5% CO₂, pH 7.4, at 30 °C. Microelectrodes were filled with either 3 M KCl or 4 M potassium acetate (resistance 50–100 M Ω). Intracellular recordings were made from pyramidal neurons in the CA3 subfield using a Dagan single microelectrode current-voltage clamp amplifier employing 2–3 kHz switching frequency and 25% duty cycle. Sampled currents and



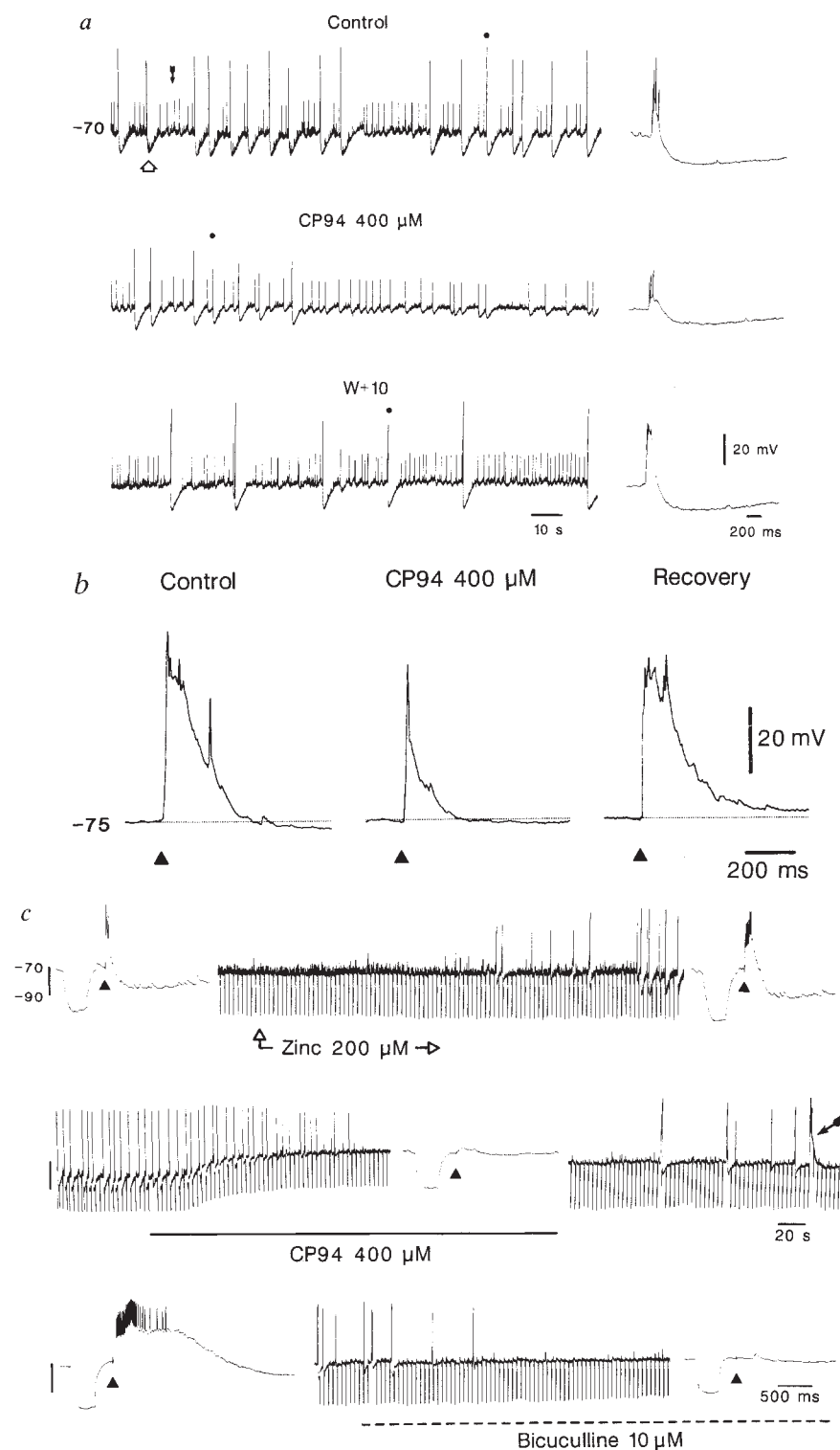
membrane voltage were recorded on a Racal store 4 tape recorder (DC to 5 kHz) and replayed onto a Brush-Gould 2200 pen recorder. Stock solutions of ZnCl₂ (Analar, BDH), CP94 and CP40 were made in distilled water and applied via the superfusate, flow rate 6 ml min⁻¹ (bath volume 1 ml). Synaptic responses were evoked in young and adult hippocampal slices by 0.1 ms, 4–8 V or 0.1 ms, 20–50 V stimuli respectively, applied to bipolar stainless steel electrodes (separation 50 μ m) located either in the hilar region of the dentate gyrus or in the stratum lucidum to stimulate the mossy fibre system. All neurons included in this study had a resting membrane potential between -60 to -70 mV with spikes > 90 mV. The double-barrelled ionophoretic electrodes were filled with 1 M GABA (pH 4, dissolved in distilled water) and 200 mM glutamate (pH 7, dissolved in 0.9% (w/v) NaCl solution). The ionophoretic pipettes were positioned in the apical dendrites of the CA3 subfield in stratum radiatum.

coefficients to either permeate the cell membrane, or to remain exclusively extracellular. Bath application of 10–400 μM CP94 (1,2-diethyl-3-hydroxypyridin-4-one, which can directly enter cells) to young CA3 neurons, resulted in a dose-dependent reduction in the amplitude and frequency of spontaneously occurring GDPs, with eventual abolition at high doses (Fig. 2a; $n = 22$). The inhibition was fully and rapidly reversible when CP94 was washed out. Spontaneous action-potential firing was unaffected by CP94, suggesting that this agent was not simply interfering with spike generation thereby blocking the production of the GDPs (Fig. 2a).

Direct orthodromic stimulation of the mossy fibre pathway

in immature slices produced an excitatory postsynaptic potential (e.p.s.p.) immediately followed by an 'evoked-GDP' which could be inhibited by either bicuculline or tetrodotoxin and also reduced in a reversible manner by CP94 (Fig. 2b). Mossy fibre stimulation will cause the release of zinc into the extracellular space^{11–13} which may be responsible for the appearance of GDPs, particularly as CP40 (1-hydroxyethyl-3-hydroxy-2-methylpyridin-4-one), a chelator which does not permeate into cells, also readily inhibited spontaneous GDPs ($n = 3$). The occurrence of GDPs gradually waned with development, such that P13–15 animals usually do not exhibit any spontaneous GDPs¹⁷, but occasionally a small evoked GDP could still be

FIG. 2 a, Zinc-chelating agent reduces the frequency of spontaneous depolarizing potentials in a young (P13) CA3 neuron. Intracellular recording of spontaneous GDPs (open arrow) which occur with sporadic action potential firing (filled arrow; top trace; note, spike amplitudes are truncated). CP94 (400 μM) reduces the amplitude and frequency of the spontaneous GDPs without affecting the action potentials (middle trace). Recovery was obtained following a 10-min wash (bottom trace). Selected GDPs from each trace (●) are shown on an expanded timescale in the righthand column. b, Averaged evoked GDPs recorded from another young (P12) CA3 neuron in the presence and absence of CP94, produced by stimulating the mossy fibre system with a 0.1 ms, 5 V stimulus (▲). Each average was constructed from five consecutively evoked GDPs using a synaptic current analysis program (SCAN, version 3.0). The dotted line indicates the resting potential, adjusted with DC current injection to -75 mV. c, Sequential recording from a P13 CA3 neuron which did not initially exhibit any spontaneous GDPs. Stimulation of the mossy fibres in Krebs' solution, before the addition of zinc, produced a small evoked GDP (top trace, ▲). Application of 200 μM Zn^{2+} throughout the experiment induced the appearance of spontaneous GDPs after 6-min incubation and prolonged the evoked GDP response (top trace). The addition of CP94 (400 μM , —), reversed the zinc-induced hyperpolarization and membrane resistance increase, and also inhibited both spontaneous and evoked GDPs (middle trace). Spontaneous GDPs recovered after wash-out of CP94 (middle trace) with zinc-containing Krebs' solution including an interictal-like discharge in this particular neuron (arrow). The evoked GDP recovered in 200 μM Zn^{2+} (bottom trace). Both spontaneous and evoked GDPs were blocked by 10 μM bicuculline (Bic) (bottom trace, - - -). The lefthand calibration indicates the membrane potential (mV) of the cell. Hyperpolarizing electrotonic potentials were applied throughout to monitor membrane resistance (-0.3 nA, 300 ms, 0.3 Hz).



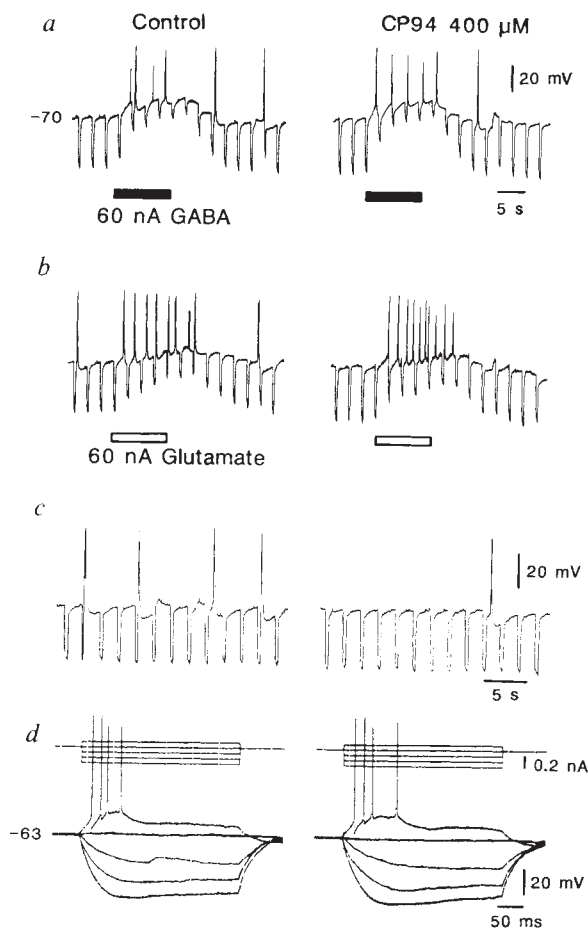
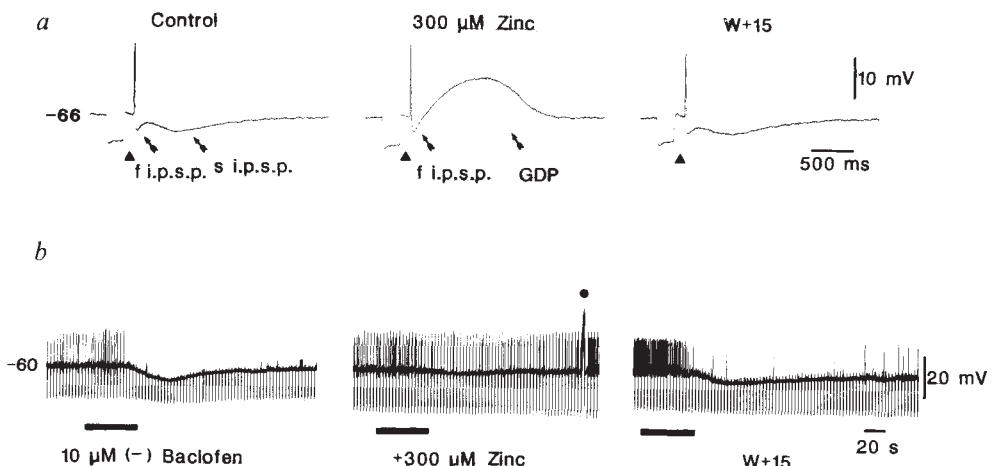


FIG. 3 Effect of CP94 on GABA and excitatory amino-acid receptors. Intracellular recording at -70 mV (with DC current injection) from a young CA3 neuron (P12) with superimposed hyperpolarizing electrotonic potentials (-0.5 nA, 300 ms, 0.5 Hz). GABA (a, solid bar) and glutamate (b, open bar) were applied for 10 s (+ and -60 nA ejection currents respectively) from a double-barrelled ionophoretic pipette to the apical dendrites. A 10 nA holding current was used on each barrel. Both GABA and glutamate produced reproducible depolarizations and conductance increases which were unaffected by $400 \mu\text{M}$ CP94. c, In the same cell, the frequency of spontaneous GDPs was reduced in the presence of CP94. d, Current clamp records of membrane potential (lower traces) obtained from another young (P12) CA3 pyramidal neuron at a resting potential of -63 mV (spike amplitude 92 mV). Superimposed electrotonic potentials produced by hyperpolarizing and depolarizing current pulses (upper traces; 300 ms, 0.25 Hz) were recorded in the presence and absence of $400 \mu\text{M}$ CP94.

FIG. 4 a, Effect of zinc on hippocampal synaptic transmission in adult slices. An e.p.s.p. followed by fast and slow i.p.s.ps (arrows) were recorded in a CA3 neuron following stimulation of the mossy fibres (0.1 ms, 50 V $\blacktriangle$). Zinc ($300 \mu\text{M}$) was applied for 5 min and depressed only the slow i.p.s.p. revealing an underlying GDP. Hyperpolarizing electrotonic potentials (-0.3 nA, 200 ms, 1 pulse per 90 s) were used to monitor membrane resistance using a 4 M potassium acetate filled microelectrode. b, Chart records of bath-applied ($-$) baclofen (solid bars) responses in the presence and absence of zinc. Downward deflections are hyperpolarizing electrotonic potentials (-0.3 nA, 300 ms, 0.3 Hz) and upward deflections are spontaneous action potentials and anode-break spikes. $\bullet$ indicates a GDP induced by zinc.



induced by mossy fibre stimulation. It was possible to induce spontaneous GDPs which reproduced the properties of innate GDPs recorded from younger neurons, by the addition of zinc (50 – $200 \mu\text{M}$; $n = 10$; Fig. 2c). The duration of the evoked GDP was also increased by zinc, but subsequent application of CP94 or bicuculline blocked both the spontaneous and evoked GDPs (Fig. 2c). The spontaneous GDPs quickly returned following wash-out of CP94 in the presence of zinc and the evoked GDP displayed an 'over-recovery'.

The generation of GDPs also relied on functional excitatory synaptic transmission, as GDPs could be partially inhibited by D(-)-2-amino-5-phosphonovaleric acid (APV, $50 \mu\text{M}$), or more effectively by 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, $10 \mu\text{M}$) suggesting the involvement of mainly α -amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) and kainate (KA) receptors (particularly as zinc blocks NMDA receptors^{14,15}). Thus the zinc-chelating agents might abolish the GDPs by blocking GABA_A receptors or ion channels, or by interfering with excitatory neurotransmission. CP94 ($400 \mu\text{M}$) did not, however, block ionophoretically-applied responses to either GABA or glutamate on CA3 neurons, despite markedly reducing the frequency of spontaneous GDPs (Fig. 3; $n = 4$). A reduction in extracellular Ca^{2+} concentration by chelation was unlikely as the chelating agents are selective for complexing zinc. (R. C. Hider, personal communication). By reducing extracellular calcium concentration (from 2 mM to 1.5 mM) to simulate complexation of Ca^{2+} by 1 mM CP94, the frequency of GDPs increased markedly, rather than decreasing as predicted if Ca^{2+} complexation had occurred. In addition, CP94 had no direct effect on the resting membrane properties or excitability of either young or adult CA3 neurons (Fig. 3d).

To elucidate the mechanism underlying GDP generation, we used the adult hippocampal brain slice in which innate GDPs are absent but can be readily induced by addition of zinc. Stimulation of the mossy fibres evoked an e.p.s.p. followed by fast and slow i.p.s.ps (Fig. 4a). The fast i.p.s.p. was mediated by GABA_A receptor activation²⁰ and the slow i.p.s.p. was mediated by GABA_B receptors²¹. CP94 ($400 \mu\text{M}$) had no obvious effect on synaptic transmission in the adult slices ($n = 5$), but 50 – $300 \mu\text{M}$ zinc caused a clear inhibition of the slow i.p.s.p. revealing a late depolarizing synaptic potential (GDP) mediated by GABA_A receptors (Fig. 4a). The initial fast i.p.s.p. was either apparently unaffected, or enhanced by zinc, eventually being masked by the onset of the much larger GDP. This indicated that zinc does not block GABA_A receptors in adult or young postnatal CA3 neurons¹⁶, but may reveal the presence of GDPs by inhibiting both pre- and postsynaptic GABA_B receptors. An inhibition of postsynaptic GABA_B responses by zinc could also be directly demonstrated in both adult and young postnatal CA3 neurons by using bath-applied baclofen, a specific GABA_B

agonist²². Zinc (50–300 μ M) reversibly inhibited the baclofen-induced hyperpolarization and conductance increase at the same concentration range which induced spontaneous GDPs in pyramidal neurons ($n=9$; Fig. 4b). These extracellular zinc levels could conceivably be achieved *in vivo* by release from synaptic vesicles, where zinc concentration has been estimated to be higher than 200–300 μ M⁴.

The suggestion that zinc has a neuromodulatory role requires the fulfillment of several stringent criteria including the identification of defined zinc-containing neuronal pathways⁴, the localization of zinc in synaptic vesicles^{7–10}, demonstration of calcium-dependent release^{11,12}, an interaction with neurotransmitter receptors/ion channels^{14–16} and the presence of uptake mechanisms^{5,6}.

Evidence of a physiological role for zinc is now provided by the demonstration of giant synaptic potentials induced by zinc in adult CNS and the inhibition of a similar potential occurring spontaneously in immature CNS neurons by chelation of endogenous zinc. The rapid onset of the zinc effect and reversible inhibition by zinc chelators, renders the involvement of zinc-containing metalloenzymes in this phenomenon unlikely²³. Previous studies of zinc chelators on synaptic transmission, have produced conflicting results. Brain zinc levels were modified only in adult animals by systemic or intracerebral infusion^{24–26} with zinc chelators, or by using dietary zinc deficiency²⁷, and transmission was monitored with extracellular recording. We have observed a more profound and consistent effect of chelation of endogenous zinc in young immature animals by recording innate GDPs. We propose that these GDPs are generated and partly revealed by zinc inhibiting both pre- and postsynaptic GABA_B receptors. This will have two consequences, by enhancing the release of GABA and increasing the amplitude of GABA-mediated synaptic potentials, and also by removing the overlying GABA_B-mediated slow i.p.s.p. which could obscure the latent GDP. The physiological function of the GDPs is probably complex; GABA_A-mediated GDPs could be generated in the dendrites where depolarizing GABA_A-mediated i.p.s.ps have previously been demonstrated and suggested to have an inhibitory role^{20,28}. The large increase in chloride conductance associated with the GDP in the dendrites, could be inhibitory by providing a 'shunt' across the membrane for excitatory inputs, but the depolarization would spread electrotonically towards the soma where some facilitation of excitation may also occur²⁸. The degree and balance of inhibition–excitation would then be expected to be closely related to the extracellular levels of zinc which may provide a tonic inhibition of GABA_B receptors. Before mature synaptogenesis^{29,30}, zinc-induced GDPs could be involved in rudimentary dendritic synaptic inhibition in this 'seizure-prone' area of the CNS. Interestingly, very young children (P4–6) can suffer from benign postnatal seizures (fifth day fits) which have been correlated with low levels of zinc in cerebrospinal fluid¹⁸. Our perception of inhibitory and excitatory transmission at immature synapses may well be altered, if spontaneous giant synaptic potentials can be generated at synapses with other zinc-containing nerve fibres throughout the young telencephalon, particularly during neuronal development. □

Received 16 November; accepted 12 December 1990.

- Danscher, G. in *The Neurobiology of Zinc* Vol. A (eds Frederickson, C. J., Howell, G. A. & Kasarskis, E. J.) 227–242 (Alan Liss, New York, 1984).
- Frederickson, C. J., Kasarskis, E. J., Ringo, D. & Frederickson, R. E. *J. Neurosci. Meth.* **20**, 91–103 (1987).
- Crawford, I. L. & Connor, J. D. *J. Neurochem.* **19**, 1451–1458 (1972).
- Frederickson, C. J. *Int. Rev. Neurobiol.* **31**, 145–238 (1989).
- Wolf, G., Schutte, M. & Romhild, W. *Neuroscience Lett.* **51**, 277–280 (1984).
- Wensink, J., Molenaar, A. J., Woroniecka, U. D. & Van Den Hamer, C. J. *J. Neurochem.* **50**, 783–789 (1988).
- Ibata, Y. & Otsuka, N. *J. Histochem. Cytochem.* **17**, 171–175 (1969).
- Perez-Clausell, J. & Danscher, G. *Brain Res.* **337**, 91–98 (1985).
- Friedman, B. & Price, J. L. *J. Comp. Neurol.* **223**, 88–109 (1984).
- Holm, I. E., Andreassen, A., Danscher, G., Perez-Clausell, J. & Nielsen, H. *Histochem.* **89**, 289–293 (1988).
- Assaf, S. Y. & Chung, S. H. *Nature* **308**, 734–736 (1984).

- Howell, G. A., Welch, M. G. & Frederickson, C. J. *Nature* **308**, 736–738 (1984).
- Charlton, G., Rovira, C., Ben-Ari, Y. & Leviel, V. *Exp. Brain Res.* **58**, 202–205 (1985).
- Peters, S., Koh, J. & Choi, D. W. *Science* **236**, 589–593 (1987).
- Westbrook, G. L. & Mayer, M. L. *Nature* **328**, 640–643 (1987).
- Smart, T. G. & Constanti, A. *Brit. J. Pharmacol.* **99**, 643–654 (1990).
- Ben-Ari, Y., Cherubini, E., Corradetti, R. & Gaiarsa, J. L. *J. Physiol.* **416**, 303–325 (1989).
- Goldberg, J. & Sheehy, E. M. *Arch. Dis. Child.* **57**, 633–635 (1982).
- Hider, R. C. *et al. Biochem. Pharmacol.* **39**, 1005–1012 (1990).
- Alger, B. E. & Nicoll, R. A. *J. Physiol.* **328**, 105–123 (1982).
- Dutar, P. & Nicoll, R. A. *Nature* **332**, 156–158 (1988).
- Hill, D. R. & Bowery, N. G. *Nature* **290**, 149–152 (1981).
- Vallee, B. L. & Galde, A. *Adv. Enzymol. Relat. Mol. Biol.* **56**, 283–430 (1984).
- Crawford, I. L. & Connor, J. D. *J. Orthomol. Psychiatry* **4**, 39–52 (1975).
- Danscher, G., Shipley, M. T. & Andersen, P. *Brain Res.* **85**, 522–526 (1975).
- Doller, H. J. & Crawford, I. L. in *The Neurobiology of Zinc* Vol. B (eds Frederickson, C. J., Howell, G. A. & Kasarskis, E. J.) 163–176 (Alan Liss, New York, 1984).
- Hesse, G. W. *Science* **205**, 1005–1007 (1979).
- Andersen, P., Dingle, R., Gjerstad, L., Langmoen, I. A. & Mosfeldt Laursen, A. *J. Physiol.* **305**, 279–296 (1980).
- McLaughlin, B. J., Wood, J. G., Saito, K., Roberts, E. & Wu, J.-Y. *Brain Res.* **85**, 355–371 (1975).
- Schwartzkroin, P. A. *Dev. Brain Res.* **2**, 469–486 (1982).

ACKNOWLEDGEMENTS. We thank R. C. Hider (Kings College) for advice and providing the zinc-chelating agents CP94 and CP40, T. Smith (British Technology Group) for enabling the use of the chelating compounds and J. Dempster (University of Strathclyde) for the provision of software (SCAN, version 3.0). This work was supported by the MRC and Wellcome Trust.

A maternally inherited superantigen encoded by a mammary tumour virus

Philippa Marrack, Elenora Kushnir & John Kappler

Howard Hughes Medical Institute, Department of Medicine, National Jewish Center for Immunology and Respiratory Medicine, Denver, Colorado 80206, USA

A COLLECTION of superantigens, molecules which in combination with class II major histocompatibility complex (MHC) engage T cells bearing particular V β chains as part of their $\alpha\beta$ receptors, have recently been described^{1,2}. The mouse self superantigen, Mls-1^a, for example, in conjunction with many MHC class II proteins, engages mouse T cells bearing V β 6, 7, 8.1 and 9, almost regardless of the sequences of the other variable components of the receptors on the T cells^{3–5}. Two types of superantigen have been identified so far: first, superantigens encoded in the mouse genome, such as Mls-1^a; second, superantigens produced by bacteria, such as the staphylococcal enterotoxins^{1,2}. Although the latter type of superantigens are in many cases known to be proteins of about 220 amino acids^{6,7}, nothing is known about the structures of any of the superantigens encoded in mouse. Here we describe the properties of a new mouse superantigen. The antigen is maternally transmitted in milk and is probably encoded by a mammary tumour virus (MTV). Given the known genetic linkage between at least one of the mouse genomic superantigens and endogenous MTV integration sites⁸, it is tempting to speculate that the superantigen described here and some of the endogenous mouse superantigens are encoded by MTVs.

TABLE 1 Low frequency of V β 14-bearing T cells in C3H/HeJ mice

Strain	V β 2 ⁺ T cells (%) ^a		V β 14 ⁺ T cells (%) ^a	
	CD4	CD8	CD4	CD8
B10.BR	8.2 $\pm$ 0.4	14.1 $\pm$ 0.4	8.1 $\pm$ 0.4	9.5 $\pm$ 0.2
BALB.K	7.3 $\pm$ 0.7	12.1 $\pm$ 1.0	7.8 $\pm$ 0.8	8.6 $\pm$ 0.4
CBA/CaJ	6.9 $\pm$ 0.4	11.7 $\pm$ 0.2	6.7 $\pm$ 0.1	8.8 $\pm$ 0.3
CBA/J	8.4 $\pm$ 0.5	16.0 $\pm$ 0.6	10.4 $\pm$ 0.6	14.7 $\pm$ 0.5
C3H/HeJ	9.1 $\pm$ 0.6	13.3 $\pm$ 0.7	2.8 $\pm$ 0.3	5.0 $\pm$ 0.3

^a T cells were isolated from the peripheral lymph nodes of 8–12-week-old mice, stained with biotinylated anti-V β 2 or anti-V β 14 antibodies followed by avidin coupled phycoerythrin and fluoresceinated anti-CD4 or anti-CD8 antibodies and analysed as previously described⁹. The cell lines secreting anti-V β 2 (B20.6) and anti-V β 14 (14–2) antibodies were gifts of B. Malissen and D. Raulet, respectively. Shown here are the means and standard errors of determinations for at least three different mice of each strain.

TABLE 2 Correlation of V β 14⁺ T-cell loss with presence of MTV in mother's milk

Strain	Milk provider	Exogenous MTV in milk	CD4 ⁺ T cells bearing V β 14 (%)*
C3H/HeJ†	B10.BR	—	7.6 ± 0.0
C3H/HeJ	C3H/HeJ	+	2.3 ± 0.2
(B10.BR × C3H/HeJ) _F 1 × C3H/HeJ	(B10.BR × C3H/HeJ) _F 1	—	6.9 ± 0.3
(B10.BR × C3H/HeJ) _F 1 × C3H/HeJ	C3H/HeJ	+	2.6 ± 0.3
C3HeB/FeJ	C3HeB/FeJ	—	9.2 ± 0.0
C3H/HeSnJ	C3H/HeSnJ	—	7.8 ± 0.3
C3H/HeOuj	C3H/HeOuj	+	3.8 ± 0.3

* T cells were isolated from the peripheral blood²⁰ or lymph nodes of the mice indicated, and analysed for percentages of CD4⁺ cells bearing V β 14 as described in the legend to Table 1. Results are the means and standard errors obtained from between two and six independent mice. Mice were of either sex, and were all 7–8 weeks old.

† Mice were removed from their C3H/HeJ mothers immediately at birth and were not permitted to suckle from their natural mothers.

During a large screen of V β expression by T cells in different strains of mice, we noticed that there were fewer T cells bearing V β 14 in C3H/HeJ animals than in other strains expressing the same MHC type, H-2^k. This was true for both CD4⁺ and CD8⁺ T cells, but the effect was more pronounced in the former population (Table 1). The phenomenon was specific for V β 14 because, as shown in Table 1, T cells bearing V β 2, for example, were at roughly equal frequency in all strains examined. Increased frequencies of T cells bearing either V β 2 or V β 14 were found in CBA/J animals, probably because expression of Mls-1^a and Mls-2/3^a and IE in these mice causes deletion of T cells bearing many other V β s.

Suspecting that the result in C3H/HeJ mice was due to the expression of a V β 14-specific superantigen and consequent V β 14⁺ T-cell deletion, we bred F₁ and backcrossed animals to examine the genetics of the superantigen. In these and all subsequent experiments the frequency of V β 14-bearing cells in the CD4⁺ T-cell population will be reported, because the phenomenon affects CD4⁺ T cells more markedly than CD8⁺ cells. To our surprise, V β 14 deletion was maternally inherited (Fig. 1). Thus (B10.BR × C3H/HeJ)_F1 animals contained V β 14-bearing CD4⁺ T cells at high frequency, whereas in (C3H/HeJ × B10.BR)_F1 animals, V β 14⁺ T cells were deleted (the mother is always named first in a mouse cross). Likewise ((B10.BR × C3H/HeJ)_F1 × C3H/HeJ) animals had high levels of V β 14-bearing CD4⁺ T cells, but in the offspring of the reciprocal

cross, (C3H/HeJ × (B10.BR × C3H/HeJ)_F1), these cells were deleted.

There are three possible explanations for such maternal effects. The genes under examination may be encoded in the mitochondrial genome⁹; they may be differentially silenced depending on whether they are inherited from the father or mother¹⁰; finally, they may be carried by some transmissible factor in milk¹¹. The first explanation is unlikely to be correct in this case because the mitochondrial DNA of C3H and B10 mice is not markedly different (K. F. Lindahl, personal communication). The backcross experiment shown in Table 2 eliminates the second possibility. Were C3H/HeJ mice to contain a genomic V β 14-specific superantigen, silent if inherited from the father, this gene would become unsilenced in the gametes of female (B10.BR × C3H/HeJ)_F1 animals. Consequently in half of the progeny of these animals, the ((B10.BR × C3H/HeJ)_F1 × C3H/HeJ) animals, the V β 14⁺ T cells should be deleted. This was not observed (Table 2).

Two types of experiment were done to test the third possibility. In the first, C3H/HeJ mice were fostered at birth to a B10.BR mother, or ((B10.BR × C3H/HeJ)_F1 × C3H/HeJ) babies were fed by a C3H/HeJ mother. C3H/HeJ babies fostered to B10.BR mothers developed into adults containing high levels of V β 14⁺ T cells, unlike their C3H/HeJ-fed relatives (Table 2). Conversely, ((B10.BR × C3H/HeJ)_F1 × C3H/HeJ) mice fed by C3H/HeJ mothers had suppressed levels of V β 14⁺ T cells whereas those fed by their own (B10.BR × C3H/HeJ)_F1 mothers had high percentages of V β 14-bearing CD4⁺ T cells. These results demonstrate that the V β 14-specific superantigen in C3H/HeJ mice is transmitted from mother to pup after birth,

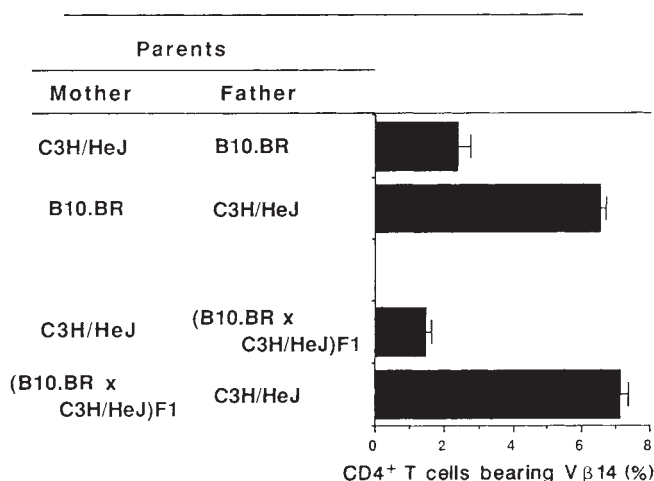


FIG. 1 V β 14 T-cell deletion is maternally inherited. B10.BR male or female mice were bred with C3H/HeJ animals of the opposite sex. (B10.BR × C3H/HeJ)_F1 male and female mice, containing V β 14⁺ T cells at high percentages, were crossed with C3H/HeJ animals of the opposite sex. For each cross the mother is indicated first. T cells were isolated from the peripheral blood or lymph nodes of the progeny and the percentages of CD4⁺ T cells bearing V β 14 in each sample were measured as described in the legend to Table 1. Results shown are the means and standard errors of determinations for at least three male and three female mice of each type. Animals were 8–12 weeks old at the time of sampling.

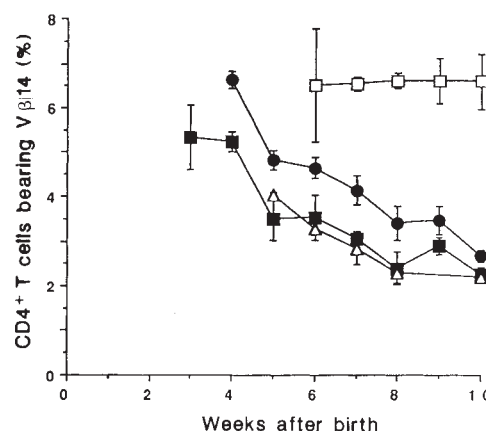


FIG. 2 V β 14-bearing T cells disappear slowly. T cells were isolated at weekly intervals from the peripheral blood of male C3H/HeJ (Δ), male (C3H/HeJ × B10.BR)_F1 (■), female (C3H/HeJ × B10.BR)_F1 (●) and male (B10.BR × C3H/HeJ)_F1 (□) mice. These cells were analysed for percentages of CD4⁺ T cells bearing V β 14 as described in the legend to Table 1. Results shown are the means and standard errors of determinations for three or four animals.

probably in milk. The only factor known to be transmitted in C3H/HeJ but not B10.BR milk is the well-known exogenous MTV characteristic of some mouse strains, including some substrains of C3H (ref. 11).

To investigate further whether exogenous MTV infection can lead to expression of the V β 14-specific deleting superantigen, four different substrains of C3H were tested for percentages of peripheral V β 14⁺ T cells. Two of these substrains (C3H/HeJ and C3H/HeOuJ) are known to express a milk-transmitted exogenous MTV. The other two substrains do not express such a virus, either because of fostering (C3HeB/FeJ) or for unknown reasons (C3H/HeSnJ)¹². The former substrains delete V β 14-bearing cells, whereas the latter do not (Table 2). These experiments support the conclusion that the V β 14-deleting superantigen is transmitted in milk, and also suggest that it may be encoded by the exogenous MTV itself. If this is so, the V β 14-specific superantigen must not be expressed by any of the endogenous MTVs known to be inserted in the genomes of C3H/HeJ or any of the other mouse strains examined here. Perhaps this reflects heterogeneity in the sequences of the exogenous and endogenous MTVs.

V β 14⁺ T cells were not deleted in young animals, and deletion occurred for a fairly long time after birth. The time course shown in Fig. 2 illustrates this point, and also demonstrates that V β 14⁺ T cells start to disappear in male mice slightly before females. This suggests that the V β 14-specific superantigen is not expressed in C3H/HeJ-related mice before birth and that its expression is controlled to some extent by sex-related hormones. The fact that these are also characteristics of milk-transmitted MTV¹³⁻¹⁵ supports the notion that the V β 14-specific superantigen described here is encoded by MTV.

Once expressed the antigen causes deletion of immature V β 14-bearing cells in the thymus (data not shown). Peripheral T-cell percentages may drop either because of turnover and dilution by new T cells bearing other V β chains and/or because of slow peripheral deletion or energy.

C3H/HeJ and nearly all other strains of mice contain several MTV integrants in their genomes. Several of these map at or close to the known genomic positions of endogenous mouse superantigens. The V β 5- and V β 11-specific superantigen, Etc-1, for example, cannot be separated in mouse strains or otherwise from MTV9⁸ (ref. 8). It is also well known that both exogenous and endogenous MTVs encode a roughly 320-amino-acid protein of unknown function in their long terminal repeats^{16,17}. Could it be that the milk-transmitted superantigen described here and some mouse self superantigens are versions of this protein?

If the superantigen described here and endogenous superantigens are encoded by MTVs or other retroviruses, what could be the advantage to the virus of such an arrangement? One explanation, a modification of a previous suggestion¹⁸, is that virally encoded superantigens might activate many T cells under the right circumstances, and these T cells might then serve as ideal targets for infection. Conversely, the mouse, whatever the source of endogenous superantigens, may use such antigens to delete T cells bearing particular V β chains and protect itself against attack by exogenous superantigens¹.

Note added in proof: Preliminary transfection experiments indicate that the MTV long terminal repeat open reading frame does indeed code for the superantigen described here (Choi, Y., J.K. and P.M., manuscript in preparation). □

10. Solter, D. A. *Rev. Genetics* **22**, 127-146 (1988).
11. Andervont, H. B. *J. natn. Cancer Inst.* **1**, 147 (1940).
12. JAX Notes 430, (1987).
13. Parks, W. P., Scolnick, E. M. & Kozikowski, E. H. *Science* **184**, 158-160 (1974).
14. Ringold, G. M. *Curr. Topics Microbiol. Immun.* **106**, 79-100 (1983).
15. Otter, A. D., Sanders, M. M. & McKnight, G. S. *Molec. Endocrinol.* **2**, 143-147 (1988).
16. Fasel, N., Pearson, K., Buetti, E. & Diggeimann, H. *EMBO J.* **1**, 3-7 (1982).
17. King, L. B., Lund, F. E., White, D. A., Sharma, S. & Corley, R. B. *J. Immun.* **144**, 3218-3227 (1990).
18. O'Neill, H. C., McGrath, M. S., Allison, J. P. & Weissman, I. L. *Cell* **49**, 143-151 (1987).
19. Kappler, J. et al. *Cell* **49**, 263-271 (1987).
20. Kappler, J. W., Kushnir, E. & Marrack, P. *J. exp. Med.* **169**, 1533-1541 (1989).

ACKNOWLEDGEMENTS. We thank K. F. Lindahl, B. Taylor and L. Schultz for their help in interpreting these experiments. This work was supported by the US Public Health Service.

Linkage of *Mls* genes to endogenous mammary tumour viruses of inbred mice

Wayne N. Frankel*, Christine Rudy†, John M. Coffin* & Brigitte T. Huber†‡

* Department of Molecular Biology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

† Department of Pathology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

T CELLS that recognize self antigen are clonally deleted in the thymus—a maturation process that occurs in the context of histocompatibility molecules and the T-cell receptor. The minor lymphocyte stimulation antigens (Mls) effect these deletions through interactions with the V β portion of the T-cell receptor, thus mimicking bacterial 'superantigens'. Intrigued by the fact that each known *Mls* gene maps to the same chromosomal region as an endogenous mouse mammary tumour virus (*Mtv*), we re-evaluated the linkage relationships between the two gene families. Here we report perfect concordance in inbred and recombinant inbred mice between the presence of four *Mtv* proviruses with the expression of *Mls* gene products. These data suggest a general model in which mammary tumour virus gene products themselves are the ligands that shape a considerable portion of the immunological repertoire of common laboratory mice.

Mtv-7, an endogenous mouse mammary tumour virus (MMTV), and *Mls-1*, which governs the deletion of T cells expressing V β 6, V β 7, V β 8.1, and V β 9 receptor molecules, have long been known to map to mouse chromosome 1 (refs 1-7). We inferred from a recent evaluation of MMTV segregation in recombinant inbred mice^{8,9} that the two loci are closely linked, with one putative recombinant in 43 recombinant inbred strains. *Mtv-7* is also present in five inbred strains containing the positive *Mls-1^a* allele (DBA/2J, SWR/J, AKR/J, SM/J and NZB/BINJ), and absent from three null *Mls-1^b* strains (C57BL/6J, C57L/J and C3H/HeJ). Given these observations, we decided to confirm and further evaluate the strain distribution of *Mtv-7*.

We first retyped AKXL and BXD recombinant inbred strains for *Mtv-7* (not shown), and found that the *Mls-1^b* AKXL-8 strain was previously misclassified as being positive for *Mtv-7* (ref. 8). Our result, which was confirmed using a cellular DNA probe to the *Mtv-7* integration site (Fig. 1), confirms the linkage and resolves the only recombinant inbred strain discordance between *Mtv-7* and *Mls-1^a*. Then, by analysing 22 inbred strains, we reproduced the *Mtv-7* typing in the aforementioned eight strains, determined that six of seven additional strains previously typed as *Mls-1^b* strains contain *Mtv-7*, and found that none of eight additional *Mls-1^b* strains contains *Mtv-7* (Fig. 2). MA/MyJ, the sole exception to the association among inbred strains, was initially classified as *Mls-1^a* because it stimulates an insulin-specific *Mls-1^a*-reactive T-cell clone¹⁰. But, unexpect-

Received 28 November; accepted 28 December 1990.

1. Marrack, P. C. & Kappler, J. W. *Science* **248**, 705-711 (1990).
2. Janeway, C. A. Jr, et al. *Immunol. Rev.* **107**, 61-88 (1989).
3. Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35-40 (1988).
4. MacDonald, H. R. et al. *Nature* **332**, 40-45 (1988).
5. Hopp, M. P., Woodland, D. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6293-6296 (1989).
6. Bettley, M. & Mekalanos, J. J. *Bact.* **170**, 34-41 (1988).
7. Jones, C. L. & Khan, S. A. *J. Bact.* **166**, 29-33 (1986).
8. Woodland, D., Hopp, M. P., Bill, J. & Palmer, E. *Science* **247**, 964-967 (1989).
9. Alzari, P. M. & Lascombe, P. R. *J. A. Rev. Immun.* **6**, 555-580 (1988).

‡ To whom correspondence should be addressed.

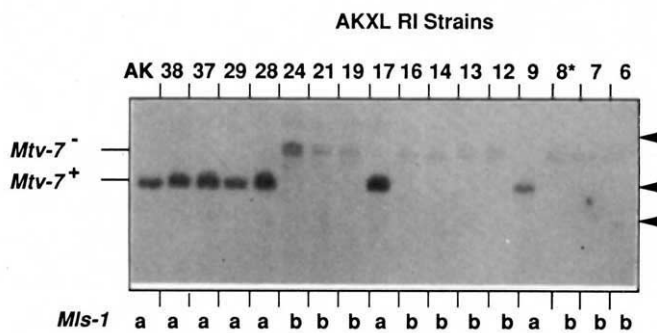


FIG. 1 Segregation analysis of *Mtv-7* in AKXL recombinant inbred (RI) strains using a flanking cellular DNA probe. Genomic DNA from AKR/J(AK), C57L/J (not shown) and AKXL RI strains were digested with *Eco*RI, Southern blotted and hybridized as previously described¹⁵. *Eco*RI fragments corresponding to the AKR/J allele (*Mtv-7*⁺) or the C57L/J allele (*Mtv-7*⁻) are shown on the left. AKXL-8 (asterisk) was previously typed as *Mtv-7*-positive⁸. Drawn below the autoradiograph is the *Mls-1* classification of these strains^{1,5,19,34}. The probe is a 4-kilobase (kb) *Hind*III-*Bgl*II fragment obtained from an *Mtv-7* junction fragment isolated from a λ library of size-selected AKXD-15 genomic DNA, and is labelled with ³²P as previously described³⁵. Numbers on the right indicate molecular size standards (kb): 23.5, 9.4 and 6.6, top to bottom.

tedly, T cells from MA/MyJ mice respond to major histocompatibility complex (MHC)-matched B cells from *Mls-1*^a strains (irrespective of other *Mls* genes) and vice versa¹¹. Furthermore, the number of V β 6, V β 7, V β 8.1 and V β 9 T cells in MA/MyJ is higher than in other *H-2*^k, *Mls-1*^a strains (C. R., E. Palmer and B. H., manuscript in preparation). This implies that the *Mls-1*^a-like phenotype of MA/MyJ mice is the product of a separate gene; in this context it is notable that MA/MyJ contains an *Mtv* fragment not present in other strains (Fig. 2). In summary, concordance between *Mls-1*^a and *Mtv-7* in 43 recombinant inbred strains shows that the two genes are not more than 2.3 centimorgans apart, using 95% confidence limits¹². Moreover, given the many opportunities for recombination between *Mls-1*^a and *Mtv-7* during inbreeding of the inbred strains, the true map distance must be even smaller.

We then reevaluated the linkage of *Mtv* proviruses with *Mls* genes governing V β 3 T-cell deletions. *Mls-3*, which was defined in DBA/2J and C3H/HeJ mice, was mapped to the middle of chromosome 16 (refs 13–15). Because *Mtv-6* resides in the same region, we retyped the AKXD and BXD RI strains for *Mtv-6*

(Fig. 3). We found that three *Mls-3*^b AKXD strains had each been previously misclassified as *Mtv-6*-positive⁸. Thus, as with *Mtv-7*/*Mls-1*, the two loci are concordant in 43 recombinant inbred strains tested.

A second determinant of V β 3 deletions in DBA/2J mice, *Mls-2*, has been mapped near *Mtv-13* on chromosome 4 (refs 13, 16). Given the other *Mtv-Mls* linkages, we were intrigued by this map location. Nevertheless, there were two putative recombinants in 29 recombinant inbred strains, and we could confirm the published recombinant inbred typing of *Mtv-13* (not shown)⁸. One discordant strain (BXD-14: *Mtv-13*-positive, *Mls-2*^b), in retrospect was not fully informative because it has the *H-2*^b haplotype and thus yields poorer T-cell deletions than *H-2*^k and *H-2*^d strains¹⁴. Interestingly, the second discordant strain (AKXD-3: *Mtv-13*-negative, *Mls-2*^a), was one of the few AKXD strains containing *Mtv-1*. Furthermore, a comparison of the extent of V β 3 deletion among 13 strains positive for *Mtv-13* reveals that the five that are *Mtv-1*-positive have significantly fewer V β 3 T cells than do the eight *Mtv-1*-negative strains; this suggests an additive deletion effect associated with the presence of *Mtv-1* (Table 1). Finally, although C3H/HeJ has the *Mls-2* phenotype, it does not contain *Mtv-13*, nor could linkage to chromosome 4 be readily demonstrated¹⁴. However, C3H/HeJ mice do contain *Mtv-1*. Thus, *Mtv-1* may represent an additional *Mls-2*-like determinant.

Taken together, these data strongly suggest a way in which a product of endogenous *Mtv* proviruses is involved in clonal deletion of specific T-cell subsets. The previously reported linkage of *Mtv-9* to an *Mls*-like gene deleting V β 5.2 T cells provides further evidence¹⁷. It is also supportive that *Mtv* proviruses are expressed in antigen-presenting B cells¹⁸ and, in particular, in *Mls-1*^a B cells (our unpublished results). One implied characteristic of the model is that some proviruses (*Mtv-7*) would express very specific epitopes recognized by the T-cell receptor, but that the V β 3 ligands are recognized as more general antigens because they are encoded by several proviruses. This is consistent with the broad distribution of the V β 3 deletion phenotype in mice, and with a pool of contributing genes¹⁹. This may reconcile how four V β 3-negative inbred strains lack *Mtv-6*, *Mtv-13* and *Mtv-1* (Fig. 2). Although these strains do not seem to share a candidate *Mtv*, their phenotype could be determined by additive effects of certain *Mtv* combinations, by a provirus whose expression varies with the genetic background, or possibly by other antigens that have the same binding properties. We

FIG. 2 Distribution of *Mtv* proviruses in 18 inbred strains of mice. *Eco*RI has often been used for identification of 5' and 3' MMTV-host DNA junction fragments^{28,9,36–38,39}, but we found that certain proviruses could more readily be scored as *Pvu*II fragments. Shown are genomic DNA from 18 inbred strains of mice, digested with *Pvu*II, blotted and hybridized to a ³²P-labelled MMTV LTR probe³⁹ as previously described³⁵. Also typed but not shown are DBA/1J, C3H/HeJ, C57BL/6J and C57L/J. For many strains, fragment identity (at the left) was inferred by *Pvu*II and *Eco*RI fragments co-migrating with those from strains analysed formally in a genetic cross (unpublished results). In most cases hybridization to a MMTV *env* probe (not shown) identified the 5' or 3' fragments. Unidentified or novel *Mtv* loci are marked with an asterisk. M; the novel *Mtv* of MA/MyJ mice. *Mtv-31* is Y-linked⁴⁰. Arrows at the left show *Mtv-1*, *Mtv-6*, *Mtv-7* or *Mtv-13* fragments. Drawn below the autoradiograph is the *Mls-1* and V β 3 classification of these strains, compiled from several sources^{1,5,7,19}. Numbers on the right are molecular size standards (kb): 23.5, 9.4, 6.6, 4.3 and 2.3, top and bottom.

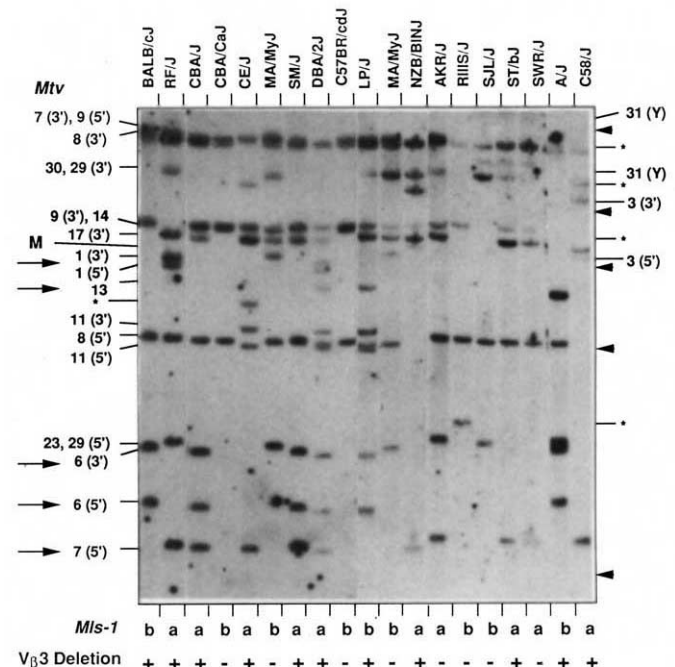
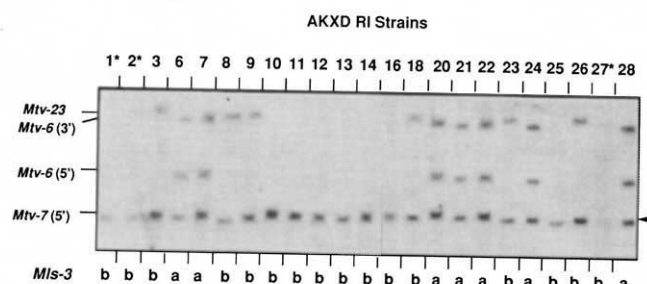


FIG. 3 Segregation analysis of *Mtv-6* junction fragments in AKXD recombinant inbred (RI) strains. *PvuII*-digested genomic DNA from AKXD RI strains was analysed on Southern blots using the MMTV LTR probe as described in Fig. 2. AKR/J and DBA/2J progenitors were shown in Fig. 2. We show only the bottom portion of the autoradiograph, to emphasize the 5' and 3' junction fragments of *Mtv-6*, as well as the 5' fragment of *Mtv-7* which does not segregate in this cross. Also present is *Mtv-23*, which comigrates with the 3' *Mtv-6* fragment. AKXD-1, -2 and -27 (asterisks) were previously scored as *Mtv-6*-positive⁸, but are clearly negative. We also scored the previously untyped AKXD-25; this strain contains *Mtv-1*, *Mtv-7*, *Mtv-8*, *Mtv-14* and *Mtv-17* (not shown). The Vβ3 phenotype of these strains is displayed below the autoradiograph, based on data from Pullen *et al.*¹³. The arrowhead at the right shows a molecular size standard of 2.3 kb.



note that *Mtv-6*, associated with the strongest Vβ3 deletion phenotype, is a defective provirus lacking the *pol* gene and possibly some of *env*²⁰. Several observations on the profound immunological effects of defective or altered retroviral gene products²¹⁻²⁵ support the possibility that some Vβ deletions are effected by novel viral gene products.

A direct test, such as proviral DNA transfection and subsequent assay for T-cell stimulation, will readily confirm these hypotheses and determine which viral gene products are responsible. Indeed, a DNA fragment containing the 3' half of *Mtv-7* (including the complete *env* gene and 3' long terminal repeat (LTR)), confers reactivity to *Mls-1*^a-negative B cells, as measured functionally *in vitro* using both Vβ6 and Vβ8.1 T-cell hybrids (U. Beutner, W.N.F., D. Smith, J.M.C. and B.T.H., unpublished results). Although the envelope glycoprotein is the obvious candidate as it is expressed on the cell surface, we cannot yet exclude the LTR open reading frame.

If *Mls* gene products are encoded by mammary tumour viruses, several interesting issues arise. First, infectious MMTV might have similar effects. Recent evidence that the milk-transmitted factor of C3H mice induces Vβ14 deletion supports this idea²⁶. Second, because MMTV amplification is associated with various T-cell lymphomas²⁷⁻²⁹, one can imagine ways in which the intrathymic selection could shift the immune system balance

in favour of frequent MMTV infection later in life. This might also be advantageous for the virus; however, it does not necessarily relate to the normal route of cell entry for MMTV, especially given the limited tissue distribution of the T-cell receptor. Finally, because *Mls* antigen mimicry of bacterial enterotoxins results in deletion of cells harmful to affected mice³⁰, it has been speculated that *Mls* genes evolved to prevent this occurrence^{31,32}. But specific MMTV proviruses closely related to known *Mtv* loci are clearly recent inhabitants of the mouse germ line, post-dating *Mus* subspeciation (<1 Myr ago), as evidenced by the limited provirus and solo LTR distribution in mice^{6-8,41}. So if *Mls* genes are *Mtv* genes, they are unlikely to have been very important to the host. Given the known mixed genealogical origins of inbred strains³³, perhaps the interaction represents a recent adaptation of an *Mtv*-naïve immune system to a novel array of antigens. □

TABLE 1 Association of *Mtv-6*, *Mtv-13* and *Mtv-1* with the Vβ3 phenotype of AKXD recombinant inbred strains

Inferred genotype	Strain	Vβ3 expression	H-2	<i>Mtv-6</i>	<i>Mtv-13</i>	<i>Mtv-1</i>
<i>Mls-2^b/Mls-3^b</i>	AKR/J	6.7 ± 0.2	k	—	—	—
<i>Mls-2^a/Mls-3^a</i>	DBA/2J	0.1 ± 0.0	d	+	+	+
<i>Mls-2^a/Mls-3^b</i>	AKXD-11	0.3 ± 0.1	k	—	+	+
	AKXD-27	0.3 ± 0.1	k	—	+	+
	AKXD-25	0.4 ± 0.1	k	—	+	+
	AKXD-1	0.8 ± 0.1	k	—	+	—
	AKXD-15	1.2 ± 0.2	k	—	+	—
	AKXD-26	1.6 ± 0.1	k	—	+	—
	AKXD-8	1.1 ± 0.2	d	—	+	+
	AKXD-9	1.1 ± 0.2	d	—	+	+
	AKXD-2	2.8 ± 0.6	d	—	+	—
	AKXD-13	2.1 ± 0.3	d	—	+	—
	AKXD-16	1.9 ± 0.2	d	—	+	—
	AKXD-18	1.9 ± 0.2	d	—	+	—
	AKXD-23	2.1 ± 0.3	d	—	+	—
<i>Mls-2²/Mls-3^a</i>	AKXD-6	0.1 ± 0.1	k	+	—	—
	AKXD-21	0.1 ± 0.1	k	+	—	—
	AKXD-24	0.4 ± 0.1	k	+	—	—
	AKXD-20	0.4 ± 0.1	k	+	—	—
<i>Mls-2^b/Mls-3^b</i>	AKXD-10	7.1 ± 0.3	k	—	—	—
	AKXD-14	6.7 ± 0.4	d	—	—	—
	AKXD-3	0.5 ± 0.1	k	—	—	+

The strain distribution of *Mtv-6*, *Mtv-13* and *Mtv-1* is shown with that of the mean (±s.e.m.) Vβ3 expression and H-2 genotypes, which were taken from Pullen *et al.*¹³. The published typings of *Mtv-1* and *Mtv-13* (ref. 8) were confirmed by us (not shown). Inferred *Mls-2* and *Mls-3* genotypes ('a' as the positive allele, 'b' as the null allele) are also shown. AKXD strains are arranged first by *Mls* genotype, then by H-2 genotype (with H-2^k being better responders than H-2^d mice), then by increasing Vβ3 expression.

Received 10 December 1990; accepted 11 January 1991.

- Festenstein, H., Bishop, C. & Taylor, B. A. *Immunogenetics* **5**, 357 (1977).
- Traina, V. L., Taylor, B. A. & Cohen, J. C. *J. Virol.* **40**, 735-744 (1981).
- MacInnes, J. I., Morris, V. L., Flintoff, W. F. & Kozak, C. A. *Virology* **132**, 12-25 (1984).
- Michalides, R., Verstraeten, R., Shen, F. W. & Hilgers, J. *Virology* **142**, 278-290 (1985).
- Abe, R., Ryan, J. J. & Hodes, R. J. *exp. Med.* **165**, 1113 (1987).
- Kozak, C. *et al. J. Virol.* **61**, 1651-1654 (1987).
- Seldin, M. F., Abe, R., Steinber, A. D., Hodes, R. J. & Morse, H. C. *J. Immunogenet.* **15**, 59 (1988).
- Lee, B. K. & Eichner, E. M. *J. Virol.* **64**, 4568-4572 (1990).
- Eicher, E. M. & Lee, B. K. *Genetics* **125**, 431-446 (1990).
- Abromson-Leeman, S. R., Laning, J. C. & Dorf, M. E. *J. Immun.* **140**, 1726 (1988).
- Ryan, J. J., LeJeune, H. B., Mond, J. J. & Finkelman, F. D. *J. Immun.* **144**, 2506 (1990).
- Silver, J. *J. Hered.* **78**, 436-440 (1985).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796-801 (1988).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *J. Immun.* **142**, 3033-3037 (1989).
- Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. *Genetics* **124**, 221-236 (1990).
- Cho, M., Villani, V. & D'Eustachio, P. *Mamm. Genome* (in press).
- Woodland, D., Happ, M. P., Bill, J. & Palmer, E. *Science* **247**, 964-967 (1990).
- King, L. B., Lund, F. E., White, D. A., Sharma, S. & Corley, R. J. *Immun.* **144**, 3218-3227 (1990).
- Abe, R., Foo-Phillips, M. & Hodes, R. J. *Immunogenetics* (in press).
- Pauley, R. J., Parks, W. P. & Popko, B. J. *Virus Res.* **1**, 381-400 (1984).
- Aziz, D. C., Hanna, Z. & Jolicoeur, P. *Nature* **338**, 505-508 (1989).
- Chattopadhyay, S. K., Morse, H. C. I., Makino, M., Ruscetti, S. K. & Hartley, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3862-3866 (1989).
- Morse, H. C., Yetter, R. A., Via, C. S. *et al. J. Immun.* **143**, 844-850 (1989).
- Li, J. P., D'Andrea, A. D., Lodish, H. F. & Baltimore, D. *Nature* **343**, 762-764 (1990).
- Yoshimura, A., D'Andrea, A. D. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4139-4143 (1990).
- Marrack, P., Kuschner, E. & Kappler, J. *Nature* **349**, 524-526 (1991).
- Michalides, R. E. *et al. J. Virol.* **43**, 819-829 (1982).
- Dudley, J. & Risser, R. J. *J. Virol.* **49**, 92-101 (1984).
- Ball, J. K., Arthur, L. O. & Dekaban, G. A. *Virology* **140**, 159-172 (1985).
- Kappler, J. W. *et al. Cold Spring Harb. Symp. quant. Biol.* **54**, 401-408 (1989).
- Marrack, P. C. & Kappler, J. W. *Science* **248**, 705-711 (1990).
- Janeway, C. A. *Cell* **63**, 659-661 (1990).
- Morse, H. in *The Mouse in Biomedical Research* (Academic, New York, 1981).
- Seldin, M. F., Morse, H. C., LeBoeuf, R. C. & Steinberg, A. D. *Genomics* **2**, 48-56 (1988).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory, New York, 1982).
- Gray, D. A., Chan, E. C. M. L., MacInnes, J. I. & Morris, V. L. *Virology* **148**, 237-242 (1986).
- Majors, J. E. & Varmus, H. E. *Nature* **289**, 253-258 (1981).
- Peters, G., *et al. J. Virol.* **59**, 535-544 (1986).
- Majors, J. E. & Varmus, H. E. *J. Virol.* **47**, 495-504 (1983).
- Prochazka, M. & Leiter, E. H. *Mouse Genome* **87**, 111 (1990).
- Stoye, J. P. & Coffin, J. M. in *RNA Tumor Viruses* (eds Weiss, N. T., Varmus, H. E. and Coffin, J. M.) 357-404 (Cold Spring Harbor, New York, 1985).

ACKNOWLEDGEMENTS. We are grateful to U. Beutner for assisting with the λ phage cloning. B. Taylor and J. Lenz for providing MMTV probes and N. Rosenberg and T. Stacy for critically reading this manuscript. We especially thank R. Hodes, P. Marrack and E. Palmer for communicating their results before publication. B.T.H. and J.M.C. were supported by the NIH. W.N.F. is the recipient of a special fellowship from The Leukemia Society of America, Inc.

An endogenous retrovirus mediating deletion of $\alpha\beta$ T cells?

David L. Woodland*[†], Mary Pat Happ[‡],
Kenneth J. Gollob[‡] & Ed Palmer[†]§||

* Department of Immunology, St Jude Children's Research Hospital,
332 N. Lauderdale, Memphis, Tennessee 38105

‡ Division of Basic Sciences, Department of Pediatrics,
National Jewish Center for Immunology and Respiratory Medicine,
1400 Jackson Street, Denver, Colorado 80206

§ Department of Microbiology and Immunology,
University of Colorado Health Sciences Center, Denver,
Colorado 80262, USA

† To whom correspondence may be addressed

|| Present address: Basel Institute for Immunology, Grenzacherstrasse 487,
CH 4005, Basel, Switzerland

A SPECIAL class of self-antigens (endogenous superantigens) is capable of deleting many murine T cells on the basis of their expression of particular T-cell receptor V β gene segments. In mice that endogenously express these antigens, tolerance is mediated in part by the clonal deletion of the relevant V β -bearing T cells¹⁻¹⁷. The deletion of I-E-reactive V β 5.2-bearing T cells is dependent on the coexpression of an I-E tolerogenic coligand (Etc)¹⁴ and the gene for one of these coligands, Etc-1, maps to chromosome 12, near the mouse mammary tumour viral integrant, Mtv-9. Here we report a perfect genetic linkage between Etc-1 and Mtv-9 and show that Etc-1 is also involved in the I-E-dependent deletion of T cells bearing V β 5.1 and V β 11 domains. We also demonstrate that Mtv-9 transcripts are present in B cells expressing Etc-1 and suggest that the coligand recognized by roughly 15% of all T lymphocytes is encoded by the Mtv-9 genome.

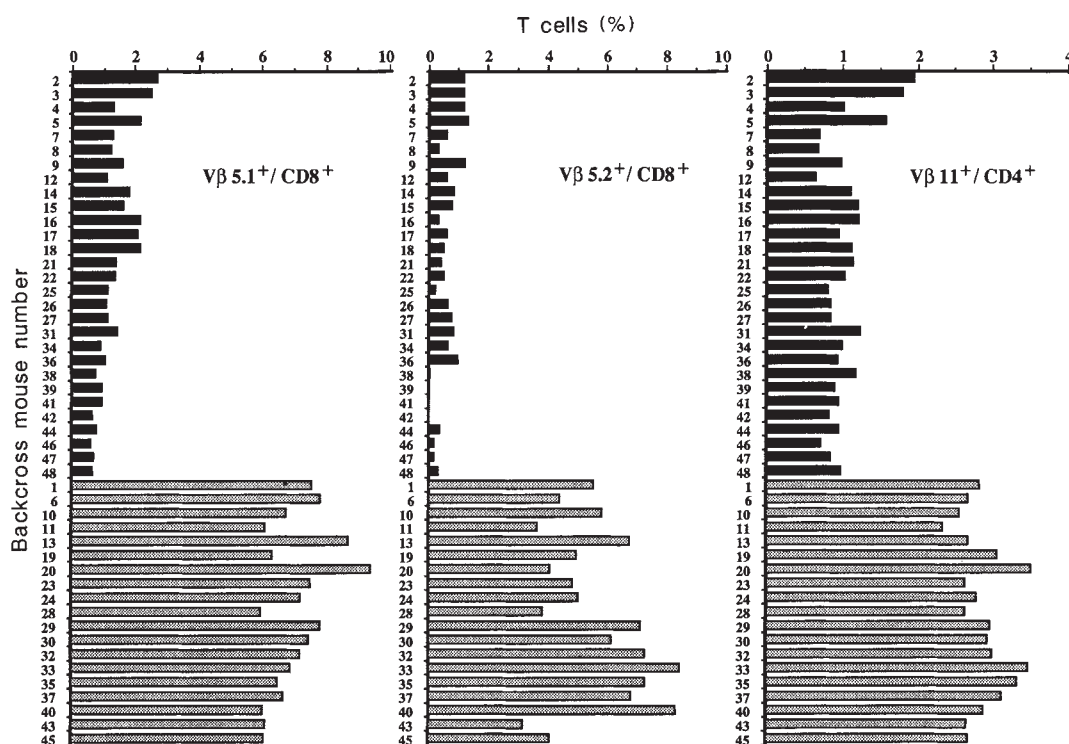
To investigate the correlation between the presence of the Mtv-9 viral genome, the expression of the Etc-1 surface antigen and the deletion of V β 5.2⁺ T cells, we extended our previous survey to include a panel of NZB $\times$ SM/J (NXSM/J) recom-

binant inbred (RI) strains^{18,19}. The results from these studies (data not shown) were consistent with our previous conclusions¹⁴ and confirmed the correlation between the presence of the Mtv-9 viral genome, the expression of the Etc-1 antigen and the strong deletion of V β 5.2-bearing T cells. In examining 37 RI strains of mice, we were unable to disassociate the phenotype of Etc-1 expression and the presence of the Mtv-9 genome. Thus, the gene encoding Etc-1 is tightly linked to the Mtv-9 locus.

To pursue the genetic linkage between Etc-1 and Mtv-9, we analysed a backcross between B10.D2 (I-E^d, Etc-1⁺, strong deleter, Mtv-9⁺) and BXD28 (I-E^d, Etc-1⁻, nondeleter, Mtv-9⁻) mice. In total, 48 (B10.D2 $\times$ BXD28)F₁ $\times$ BXD28 backcross animals were analysed for their expression of I-E/Etc-1 reactive V β 5⁺ and V β 11⁺ T cells, and Fig. 1 shows the frequencies of V β 5.1⁺/CD8⁺, V β 5.2⁺/CD8⁺ and V β 11⁺/CD4⁺ T cells in each of the 48 backcross mice. The backcross animals fall into two groups with respect to the level of V β 5⁺ and V β 11⁺ T cells (Fig. 1). Twenty-nine of 48 animals (60%) expressed low frequencies of V β 5.1⁺/CD8⁺, V β 5.2⁺/CD8⁺ and V β 11⁺/CD4⁺ T cells, similar to the B10.D2 parent whereas 19 of 48 animals (40%) expressed higher frequencies of these T cells, similar to the BXD28 parent. This pattern is consistent with the inheritance of a single gene from the B10.D2 parent that controls the elimination of I-E/Etc reactive T cells. Since there should also be a segregation of the Mtv-9 integrant in this particular cross, the 48 backcross animals were genotyped for the presence or absence of this provirus^{20,21}. There is a perfect correlation between those animals inheriting the Mtv-9 provirus (Fig. 1, solid bars) and those animals expressing low frequencies of V β 5.1⁺/CD8⁺, V β 5.2⁺/CD8⁺ and V β 11⁺/CD4⁺ T cells (strong deletion). Conversely, Mtv-9-negative mice (Fig. 1, shaded bars) all expressed higher frequencies of these T cells (weak/no deletion).

The inability to disassociate the phenotypes of Etc-1 expression or V β 5.1, V β 5.2 and V β 11 deletion from the presence of the Mtv-9 provirus led us to speculate that Etc-1 is encoded

FIG. 1 V β 5.1, V β 5.2 and V β 11 expression in 48 [(BXD28 $\times$ B10.D2)F₁ $\times$ BXD28] backcross animals. The solid bars represent V β expression in those animals which had inherited the Mtv-9 provirus and the shaded bars represent V β expression in those animals which had not inherited the Mtv-9 provirus. The data show that Mtv-9⁺ backcross mice delete V β 5.1⁺ and V β 11⁺ as well as V β 5.2⁺ T cells. METHODS. The expression of V β 5.1⁺, 5.2⁺ and 11⁺ cells was determined as previously described using the monoclonal antibodies MR9-4 (specific for V β 5.1 and V β 5.2)¹, MR9-8 (specific for V β 5.1 only)¹, RR3-15 (specific for V β 11)², 53.6 (specific for CD8)²⁶ and GK1.5 (specific for CD4)²⁷. The presence or absence of the Mtv-9 provirus was determined by restriction fragment length polymorphism analysis of EcoRI-digested DNA from each animal using a 2-kb genomic probe specific for a region of chromosome 12 which flanks the Mtv-9 viral insertion point, as described in ref. 20.



by the viral genome itself. To test whether the Mtv-9 genome is transcriptionally active in B-cell lymphomas expressing Etc-1, we developed a polymerase chain reaction (PCR) assay which takes advantage of the *Hpa*I site in the Mtv-9 envelope gene that distinguishes it from the envelope genes in three other MMTV integrants (Mtv-6, Mtv-8 and Mtv-17) in the C57BL and BALB/c genetic backgrounds^{20,21}. The amplification of envelope RNA yielded a 750-base pair (bp) fragment which could be cleaved into 300-bp and 450-bp fragments upon digestion with *Hpa*I if the template RNA had been transcribed from the Mtv-9 genome. Using this assay, we determined that Mtv-9 envelope transcripts are absent or weakly expressed in three I-E⁺ B-cell lymphomas which fail to stimulate the V β 11⁺,

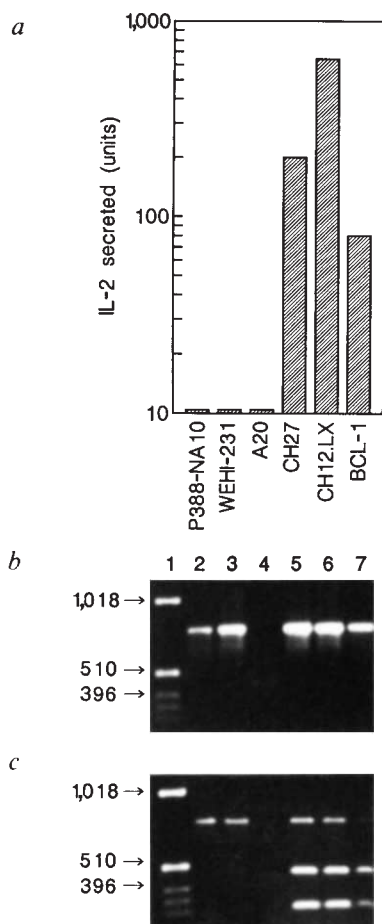


FIG. 2 **a**, IL-2 secretion by the I-E/Etc-1 reactive T-cell hybridoma 11.40 in response to several I-E⁺B-cell lymphoma lines. **b** and **c**, Transcription of the Mtv-9 proviral genome in several I-E⁺B-cell lymphoma lines. A portion of the MMTV envelope gene was amplified from RNA using the PCR generating a 750-bp fragment (**b**). The Mtv-9 genome contains a *Hpa*I site, not shared by any other Mtv integrant in BALB/c or B10 (refs 20, 21, 25), which allows a 750-bp PCR product specifically derived from Mtv-9 transcripts to be cut into 300-bp and 450-bp fragments (**c**). Mtv-9 transcription correlates with the expression of Etc-1 in B-cell lymphomas.

METHODS. The IL-2 assay was done as in ref. 14, except that only 1×10^5 B lymphomas (stimulators) were present in each well and that they were not pretreated with LPS and IL-4. RNA was prepared from the lymphoma cell lines by the method of Kumagai *et al.*²⁸. Reverse transcription and PCR were done as previously described²⁹ using the sense strand oligonucleotide AGACGAGTCTGCTCTCCAC (5' to 3') and the antisense strand oligonucleotide CGGATAAGGCTAATGTACATTTTC (5' to 3'). We then digested 10 μ l of the PCR reaction with *Hpa*I in a final volume of 100 μ l, precipitated and run on a 1.2% agarose gel. Lane 1, DNA markers; lane 2, P388-NA10; lane 3, WEHI-231; lane 4, A20; lane 5, CH27; lane 6, CH12.LX; lane 7, BCL-1. The lymphomas are derived from the following mouse strains: P388-NA10 (DBA/2, Mtv9⁻); WEHI-231 (BALB/c $\times$ NZB, Mtv9⁺); A20 (BALB/c, Mtv9⁺); CH27 (B10.H-2^aH-4^bWts, Mtv9⁺); CH12.LX (B10.H-2^aH-4^bWts, Mtv9⁺); BCL-1 (BALB/c, Mtv9⁺)³⁰.

Etc-1-reactive hybrid, 11.40. On the other hand, in three I-E⁺ B-cell lymphomas expressing Etc-1, Mtv-9 envelope transcripts are readily detected. Thus, Etc-1 expression correlates with the transcription of the Mtv-9 provirus (Fig. 2).

We have studied a member of the Etc family of antigens, Etc-1. Using two independently derived T-cell hybridomas and the deletion patterns of V β 5.1⁺, V β 5.2⁺ and V β 11⁺ T cells, we were able to show a perfect concordance between the inheritance of Etc-1 and the Mtv-9 provirus. That we could not find a single example of genetic recombination segregating these two loci when examining 37 RI strains and 48 backcross animals, indicates that these two genes are linked on a segment of chromosome 12 spanning 0.52 ± 0.53 centimorgans (cM), representing approximately 800 kb²². The linkage between Etc-1 and Mtv-9 on chromosome 12 suggested that Etc-1 is actually encoded by the Mtv-9 genome itself, an idea which is supported by several observations. In the spleen, Etc-1 is primarily expressed on B lymphocytes and its expression is greatly enhanced when B cells are treated with the mitogen, lipopolysaccharide (LPS) and the cytokine, interleukin (IL)-4 (K.G. and E.P., unpublished data). Corley and coworkers have found Mtv-9 proviral transcripts in normal B cells and B-cell tumours derived from Mtv-9⁺ mice^{23,24}. Furthermore, LPS treatment increases the level of Mtv-9 transcription 10–20-fold²⁵. In terms of their tissue expression and inducibility by mitogenic agents, the Mtv-9 genome and the Etc-1 antigen are similar. Using a PCR assay that specifically detects Mtv-9 transcripts, we showed that Mtv-9 envelope transcripts are absent or weakly expressed in three Etc-1-negative B-cell lymphomas and readily detected in three Etc-1-positive B-cell lines. Thus, the transcriptional activity of the Mtv-9 provirus correlates with expression of the Etc-1 antigen.

Our findings are consistent with those of Frankel *et al.* (manuscript submitted), Marrack *et al.* (manuscript submitted) and Tomonari *et al.* (manuscript submitted). That a large fraction of T cells react with an MMTV gene product is intriguing. This may be advantageous to the infectious virus if an interaction with T lymphocytes is an important step in the viral life cycle. Furthermore, there may be a selective advantage to mice that express the noninfectious, endogenous MMTVs if deletion of the relevant V β -bearing T cells confers a resistance to infection with exogenous mouse mammary tumour viruses. □

Received 29 November 1990; accepted 16 January 1991.

- Bill, J., Kanagawa, O., Linten, J., Utsunomiya, Y. & Palmer, E. *J. molec. cell. Immunol.* **4**, 269–280 (1990).
- Bill, J., Kanagawa, O., Woodland, D. & Palmer, E. *J. exp. Med.* **169**, 1405–1419 (1989).
- Tomonari, K. & Lovering, E. *Immunogenetics* **28**, 445 (1988).
- Okada, C. Y. & Weissman, I. L. *J. exp. Med.* **169**, 1703–1719 (1989).
- Bill, J., Appel, V. B. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9184–9188 (1988).
- Kappler, J. W. *et al. Cell* **49**, 263–272 (1987).
- Kappler, J., Roehm, N. & Marrack, P. *Cell* **49**, 273–280 (1987).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796–801 (1988).
- MacDonald, H. R. *et al. Nature* **332**, 40–45 (1988).
- Kanagawa, O., Palmer, E. & Bill, J. *J. Cell Immunol.* **119**, 412–426 (1989).
- Okada, C. Y., Holzmann, B., Guidos, C., Palmer, E. & Weissman, I. L. *J. Immunol.* **144**, 3473–3477 (1990).
- Kappler, J., Staerz, U., White, J. & Marrack, P. *Nature* **332**, 35–40 (1988).
- Happ, M. P., Woodland, D. L. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6293–6296 (1989).
- Woodland, D., Happ, M. P., Bill, J. & Palmer, E. *Science* **247**, 964–967 (1990).
- Vacchio, M. S., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **172**, 802–813 (1990).
- Tomonari, K. *Immunol. Rev.* **116**, 139–157 (1990).
- Palmer, E., Woodland, D. L., Happ, M. P., Bill, J. & Kanagawa, O. *Cold Spring Harb. Symp. quant. Biol.* **LIV**, 135–145 (1989).
- Eicher, E. M. & Lee, B. K. *Genetics* **125**, 431–446 (1990).
- Lee, B. K. & Eicher, E. M. *J. Virol.* **64**, 4568–4572 (1990).
- Peters, G. *et al. Virology* **59**, 535–544 (1986).
- Kozak, C. *et al. J. Virol.* **61**, 1651–1654 (1987).
- Taylor, B. A. in *Origin of Inbred Mice* (ed. Morse, H. C.) 423–438 (Academic Press, New York, 1987).
- King, L. B. & Corley, R. B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2814–2819 (1989).
- Sharma, S., King, L. B. & Corley, R. B. *J. Immunol.* **141**, 2510–2518 (1988).
- King, L. B., Lund, F. E., White, D. A., Sharma, S. & Corley, R. B. *J. Immunol.* **144**, 3218–3227 (1990).
- Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **47**, 63–90 (1979).
- Dialynas, D. *et al. J. Immunol.* **131**, 2445–2451 (1983).
- Kumagai, N., Benedict, S., Mills, G. & Gelfand, E. *J. Immunol.* **139**, 1393–1399 (1987).
- Bill, J. *et al. J. exp. Med.* **169**, 115–133 (1989).
- Pennell, C. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3799–3803 (1985).

ACKNOWLEDGEMENTS. We thank E. Eicher for providing the NXSM/J RI strains, C. Dickson and D. Peterson for providing DNA probes, D. Cohen, A. Finnegan and J. Cambier for providing cell lines, C. Steinberg and R. Corley for discussion and M. Cohn for a review of the manuscript. This work is supported by the NIH. E.P. is a recipient of a Faculty Research Award from the American Cancer Society.

Genes encoding ligands for deletion of $V\beta 11$ T cells cosegregate with mammary tumour virus genomes

P. J. Dyson, A. M. Knight, S. Fairchild, E. Simpson & K. Tomonari

Transplantation Biology, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

THE T-cell receptor (TCR) repertoire is selected in the thymus after rearrangement of genes encoding TCR α and β chains¹. Selection is based on the recognition by newly emergent T cells of self-ligands associated with molecules of the major histocompatibility complex: some combinations result in positive selection, others in negative selection. Negative selection, or clonal deletion, is an important mechanism for eliminating autoreactive T cells. A group of self-ligands involved in clonal deletion was identified because they, like exogenous superantigens², were recognized by almost all T cells expressing particular TCR $V\beta$ genes. $V\beta 17a$ T cells are deleted by a tissue-specific ligand^{3,4}; $V\beta 6$, $V\beta 7$, $V\beta 8.1$ and $V\beta 9$ T cells are deleted by the minor lymphocyte-stimulating (Mls) determinant Mls-1^a (refs 5–8); $V\beta 3$ T cells by Mls-2^a and Mls-3^a (refs 9, 10); $V\beta 11$ T cells¹¹ by ligands encoded by independently segregating genes; and $V\beta 5$ T cells by ligands encoded by two genes¹². Chromosome mapping using recombinant inbred strains of mice and classic backcrosses show that Mls-1^a in DBA/2 mice is encoded on chromosome 1, that one of the two ligand genes for deletion of $V\beta 5$ T cells maps to chromosome 12 (ref. 12) and that a ligand gene for $V\beta 11$ deletion is linked to the CD8 locus on chromosome 6 (ref. 11). Here we present evidence from three sets of backcross mice for concordance between $V\beta 11$ deletion ligand genes on chromosomes 6, 12 and 14 and endogenous mouse mammary tumour virus integrant (Mtv) genomes. Our results indicate that the $V\beta 11$ deletion ligands are products of Mtv genomes.

The product of the Dvb11-1 locus, in association with I-E, leads to partial deletion of $V\beta 11^+$ T cells in several mouse strains and F_1 hybrids^{11,13,14}. This locus has been separated from CD8 (mouse chromosome 6) by two recombination events in 257 animals deriving from seven backcross combinations¹¹. Typing of these two recombinants with two further markers closely linked to CD8, a polymorphic microsatellite from the Ig κ region¹⁵ and Mtv-8 probes¹⁶ revealed nonrecombinant genotypes (data not shown). Similarly, typing of 87 animals from the nondeleted and partially deleted classes of three backcrosses revealed no recombination between Dvb11-1 and these two markers (Table 1).

Other mice from these backcrosses show a total deletion phenotype mediated by the action of an independently segregating gene, Dvb11-2 (ref. 11). Using polymorphic microsatellites¹⁷ to type the (B10.BR $\times$ C58) F_1 $\times$ C58 backcross, a weak linkage (22% recombination, data not shown) was found between the immunoglobulin heavy chain region (Igh-V), close to the telomere of chromosome 12, and Dvb11-2. This analysis places Dvb11-2 near Mtv-9. Typing of this backcross for Mtv-9 using the mouse mammary tumour virus (MMTV) 3' probe¹⁸ revealed 1 discordant animal out of 71 (Table 1). This discordant animal (see below) was in the totally deleted class for $V\beta 11^+$ T cells but was negative for Mtv-9; it did, however, possess Mtv-8. Typing of 52 animals of the (BALB/c $\times$ C58) F_1 $\times$ C58 backcross for Mtv-9 revealed no animals with discordant Mtv-9 genotype and Dvb11-2 phenotype (Table 1). When the Igh-V microsatellite was used to type the (C3H $\times$ C58) F_1 $\times$ C58 backcross, surprisingly no linkage (data not shown) with the gene causing almost total deletion in C3H was observed. The observation that the gene encoding the totally deleting ligand must map to another location in C3H as compared with B10.BR and BALB/c is intriguing and prompted us to search for its location. Because linkage was observed to endogenous Mtv for two genes responsible for deleting $V\beta 11^+$ T cells (Dvb11-1, Dvb11-2), the (C3H $\times$ C58) F_1 $\times$ C58 backcross was typed for Mtv using the MMTV 3' probe. In the 74 animals analysed (Table 1), no discordance was observed between Mtv-11 (chromosome 14) and the $V\beta 11^+$ T cell complete-deletion phenotype (Table 1 and Fig. 1).

The data presented here extend earlier observations on the genes that delete $V\beta 11^+$ T cells^{11,13,14,19–21}. We propose that the genes responsible for total deletion be named Dvb11-2 (chromosome 12) and Dvb11-3 (chromosome 14). The most striking observation is the cosegregation of Dvb11-1, -2 and -3 with Mtv-8, -9 and -11, respectively. Table 2 illustrates that for the six strains typed for $V\beta 11^+$ T cell deletion and Mtv-8, -9 and -11, a correlation is seen between the presence of Mtv-8, -9 and -11 and Dvb11-1, -2 and -3, respectively.

The penetrance of Dvb11-1 is variable; this is manifest in the range of partial deletion seen in a variety of backcrosses¹¹. Further, in the cross (C58 $\times$ SJL) F_1 $\times$ C58, Dvb11-1 leads to a complete deletion phenotype²¹ as it did in a single second-backcross-generation animal from the (CBAT6 $\times$ C58) F_1 $\times$ C58 cross¹¹. The one discordant animal in the (B10.BR $\times$ C58) F_1 $\times$ C58 backcross was phenotypically Dvb11-2⁺, but had the genotype Mtv-8⁺, Mtv-9[–]; the total deletion phenotype is likely to be due to the action of Dvb11-1 which, in this backcross, leads to a particularly wide continuum of partial deletion¹¹ which may overlap the total deletion caused by Dvb11-2. Consistent with this, the frequency of $V\beta 11^+$ T cells is 1.9%, placing them at the upper end of the total deletion class. It will be of interest to establish the basis for the variable penetrance of Dvb11-1 in different genetic backgrounds.

FIG. 1 Southern blot analysis illustrating the segregation of Mtv genomes in (C3H $\times$ C58) F_1 $\times$ C58 backcross mice. Left panel, two parental strains with the Mtv genomes indicated (the two Mtv genomes present in C58 in addition to Mtv-17 have not been identified³²). Middle and right panel, 10 animals from the nondeleted and totally deleted backcross classes respectively. Methods were as described in Table 1 legend. Size markers are the *Hind* III fragments of λ bacteriophage.

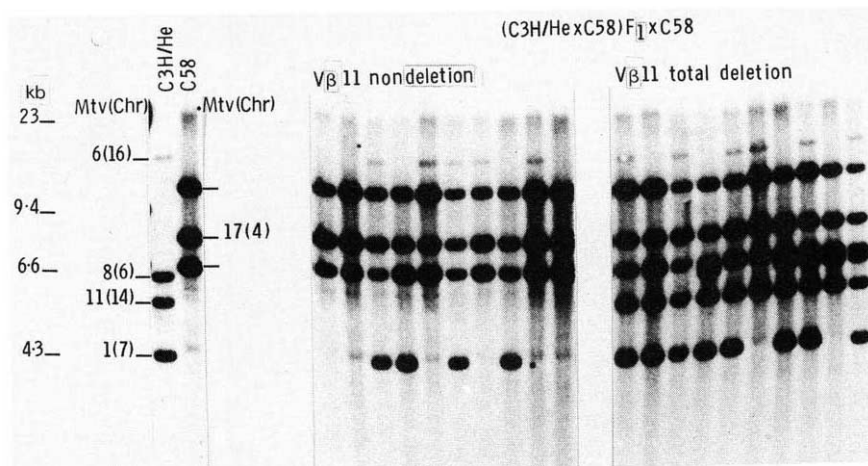


TABLE 1 Inheritance of genetic markers on chromosomes 6, 12 and 14 in backcross mice showing the three V β 11 deletion phenotypes

Backcross	V β 11 deletion phenotype	Range	Proposed name	Locus (chromosome)				
				CD8(6)	Ig κ (6)	Mtv-8(6)	Mtv-9(12)	Mtv-11(14)
(B10.BR $\times$ C58)F ₁ $\times$ C58 (73 mice)	Partial	(2.3%–5.9%)	Dvb11-1	¹⁵ /15	¹⁵ /15	¹⁴ /14	⁰ /15	NP
	Total	(0.3%–1.9%)	Dvb11-2	¹⁸ /37	¹⁸ /37	¹⁷ /36	³⁵ /36*	NP
	Non	(9.4%–11.3%)		⁰ /21	⁰ /21	⁰ /20	⁰ /20	NP
BALB/c $\times$ C58)F ₁ $\times$ C58 (52 mice)	Partial	(2.3%–6.1%)	Dvb11-1	⁷ /7	⁷ /7	⁷ /7	⁰ /7	NP
	Total	(0.2%–1.4%)	Dvb11-2	¹⁸ /30	¹⁸ /30	¹⁸ /30	³⁰ /30	NP
	Non	(9.5%–10.5%)		⁰ /15	⁰ /15	⁰ /15	⁰ /15	NP
(C3H/He $\times$ C58)F ₁ $\times$ C58 (74 mice)	Partial	(4.5%–7.5%)	Dvb11-1	¹⁵ /15	¹⁵ /15	¹⁵ /15	NP	⁰ /15
	Total	(1.0%–2.2%)	Dvb11-1	²³ /43	²³ /43	²³ /43	NP	⁴³ /43
	Non	(9.4%–11.1%)		¹ /16	⁰ /16	⁰ /16	NP	⁰ /16

Mice were bred and maintained at the Clinical Research Centre. Phenotyping of T-cell subpopulations was performed as described¹¹ with lymph-node cells being strained with biotinylated KT11 (anti-TCR V β 11) and analysed on a fluorescence-activated cell sorter (FACStar). DNA was prepared from each backcross mouse as described²⁹. Genetic markers MUSLYT3A1 (CD8) and MMIGK (Ig κ) are microsatellite repeats. Primers and conditions for CD8 were as described¹⁷. To amplify a (TG)₂₀ repeat in the Ig κ (ref. 30) locus, 5' GAT TCC CTC TTC TGG TGT ATG and 3' GTC ATT ACA TGT ATT CTC TGT primers were used. Amplification conditions: 32 cycles of a reaction at 94 °C for 1 min, 55 °C for 1 min and 72 °C for 30 s with 5 mM Mg²⁺. All PCR products were analysed nonradioactively with 4% Nuseive agarose and ethidium-bromide staining. MMTV provirus inheritance patterns were followed using Southern blotting techniques²⁹. Probes used were the cloned milk-borne MMTV 3' probe and the specific Mtv-8 3' flanking probe (Mtv8-3'C)¹⁸. DNA (10 μ g) was restricted with EcoRI (BRL) and following size separation through 0.7% agarose transferred to nylon membranes (Genescreen, Dupont). Probes described were labelled with random oligo-primed [α -³²P]dCTP to 10⁸ c.p.m. After overnight hybridization at 65 °C, filters were washed at high stringency (0.1 $\times$ SSC, 0.1% SDS) for 2 h at 65 °C. Filters were dried and exposed overnight with X-ray film (Kodak) and intensifying screens. Numbers indicate numbers of mice inheriting alleles from the deleting strain out of the total tested. NP, viral marker not present. Levels of V β 11⁺ T cells in parental strains were as follows: C58 (10.8 $\pm$ 0.4%), B10.BR (2.3 $\pm$ 0.2%), BALB/c (1.3 $\pm$ 0.2%), C3H/He (0.9 $\pm$ 0.1%) (data from ref. 19).

* See text.

Dvb11-1, -2 and -3 share one important feature with the minor lymphocyte-stimulating loci Mls-1^a, -2^a and -3^a, and exogenous superantigens—the ability to delete part or a whole class of T cells expressing a particular TCR V β gene(s)²². The gene encoding a ligand responsible, in association with I-E, for the deletion of V β 5⁺ T cells has been mapped near to Mtv-9 (ref. 12). It is not so far known whether there is any relationship between this gene and Dvb11-2. Our data are consistent with Mtv-8, -9 and -11 being very closely linked to, or identical with, Dvb11-1, -2 and -3. The products encoded by Mtv include polypeptides associated with the plasma membrane which could interact with class II molecules, a property shown by bacterial toxin superantigens^{23,24} and proposed for Mls products²⁵. Although Mtv is principally expressed under hormonal regulation in lactating mammary tissue, expression in lymphoid tissue is documented^{26,27}. For Mtv to be responsible for deletion of T cells, presence of viral products on a thymic stromal cell type would be expected; this is now under investigation. If Mtv genomes encode ligands that can negatively select T cells expressing particular V β TCR genes, their evolutionary persistence may be related to beneficial changes in the T-cell repertoire (perhaps by excluding the maturation of T cells reactive with particular exogenous bacterial toxins, as proposed for Mls loci²²). It could also relate to an advantage conferred by the elimination of potentially autoreactive T-cell clones²⁸.

In the backcrosses described here, several Mtv genomes are segregating in addition to Mtv-8, -9 and -11 (Fig. 1 and data not shown); no cosegregation was observed for three proviruses and the V β 11⁺ T-cell deletion phenotypes. It will be interesting to ascertain whether these viral genomes mediate T-cell selection events involving other TCR V β genes. We are directly testing the hypothesis that Mtv gene products can function as co-ligands for the deletion of particular V β ⁺ T cells by generating transgenic mouse lines and transfectant cell lines carrying cloned Mtv genomes. □

Received 13 December 1990; accepted 22 January 1991.

- Lawlor, D. A., Zemmour, J., Ennis, P. D. & Parham, P. A. *Rev. Immun.* **8**, 23–63 (1989).
- White, J. et al. *Cell* **56**, 27–35 (1989).
- Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273–280 (1987).
- Marrack, P. & Kappler, J. *Nature* **332**, 840–843 (1988).
- Kappler, J. W., Staerz, U., White, J. & Nature **332**, 35–40 (1988).
- MacDonald, H. R. et al. *Nature* **332**, 40–45 (1988).
- Okada, C. Y., Holzmann, B., Guidos, C., Palmer, E. & Weissman, I. L. *J. Immun.* **144**, 3473–3477 (1990).
- Happ, M. P., Woodland, D. C. & Palmer, E. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6293–6296 (1989).
- Abe, R., Foo-Phillips, M. & Hodes, R. J. *J. exp. Med.* **170**, 1059–1073 (1989).
- Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335**, 796–801 (1988).
- Tomonari, K. & Fairchild, S. *Immunogenetics* (in the press).
- Woodland, D., Happ, M. P., Bill, J. & Palmer, E. *Science* **247**, 964–967 (1990).
- Tomonari, K. & Lovering, E. *Immunogenetics* **28**, 445–451 (1988).
- Bill, J., Kanagawa, O., Woodland, D. L. & Palmer, E. *J. exp. Med.* **169**, 1405–1419 (1989).
- Gibson, D. M., Maclean, S. J. & Cherry, M. *Immunogenetics* **18**, 111–116 (1983).
- Davidsson, M. T. & Roderick, T. H. in *Genetic Variants and Strains of the Laboratory Mouse* 2nd edn (eds Lyon, M. F. & Searle, A. G.) 416–427 (Oxford University Press, Oxford, 1989).
- Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. *Nucleic Acids Res.* **18**, 4123–4130 (1990).
- Peters, G. *J. Virol.* **59**, 535–544 (1986).
- Tomonari, K. *Immunol. Rev.* **116**, 139–157 (1990).
- Vacchio, M. S., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **172**, 807–813 (1990).
- Tomonari, K. *Immunogenetics* **32**, 60–62 (1990).
- Marrack, P. & Kappler, J. *Science* **248**, 705–711 (1990).
- Fraser, J. D. *Nature* **339**, 221–224 (1989).
- Mollick, J. A., Coote, R. G. & Rich, R. R. *Science* **244**, 817–821 (1989).
- Janeway, C. A. Jr. *Immunol. Rev.* **107**, 61–88 (1989).
- Henrard, D. & Ross, R. R. *J. Virol.* **62**, 3046–3049 (1988).
- King, L. B., Lund, F. E., White, D. A., Sharma, S. & Corley, R. B. *J. Immun.* **144**, 3218–3227 (1989).
- Heber-Katz, E. & Acha-Orbea, H. *Immun. Today* **10**, 164–169 (1989).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Gebhard, W. & Zachau, H. G. *J. molec. Biol.* **170**, 567–573 (1983).
- Knight, A. M. & Dyson, P. J. *Molec. cell. Probes* **4**, 497–504 (1990).
- Yang, J.-N., Boyd, R. T., Gottlieb, P. D. & Dudley, J. P. *Immunogenetics* **25**, 222–227 (1987).

ACKNOWLEDGEMENTS. We thank G. Peters for the Mtv probes and discussions, J. Todd for support, D. B. Palmer for suggestions and V. Tikerpa for assistance in preparing the manuscript. This work was supported in part by the Cancer Research Campaign.

TABLE 2 Correlation of Mtv insertions with the presence of Dvb11-1, Dvb11-2 and Dvb11-3 in inbred strains

Strain	MMTV present			Deletion genes		
	Mtv-8	Mtv-9	Mtv-11	Dvb11-1	Dvb11-2	Dvb11-3
C58	—	—	—	—	—	—
NOD	—	—	—	—	—	—
B10.BR	+	+	—	+	+	—
BALB/c	+	+	—	+	+	—
C3H/He	+	—	+	+	—	+
A	+	—	—	+	—	—

Correlation between presence of Mtv-8, -9, -11 and Dvb11-1, -2, -3 respectively. —, not present; +, present. Data from refs 11, 18 and 31 and this study.

Induction of CD4 and susceptibility to HIV-1 infection in human CD8⁺ T lymphocytes by human herpesvirus 6

Paolo Lusso*, Andrea De Maria†, Mauro Malnati‡, Franco Lori*, Susan E. DeRocco§, Michael Baseler|| & Robert C. Gallo*

* Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, Maryland 20892, USA

† Laboratory of Immunoregulation and ‡ Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892, USA

§ Department of Cell Biology, Advanced BioScience Laboratories, Kensington, Maryland 20895, USA

|| Program Resources Inc., National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

DURING intrathymic T-cell ontogenesis, functionally competent CD3⁺CD4⁺CD8⁻ and CD3⁺CD4⁻CD8⁺ T lymphocytes develop from immature CD4⁻CD8⁻ thymocytes after transiently acquiring a double-positive CD4⁺CD8⁺ phenotype¹⁻³. The partition between CD4⁺CD8⁻ and CD4⁻CD8⁺ T cells is generally considered to be irreversible, although a small percentage of circulating CD3⁺ T lymphocytes coexpressing CD4 and CD8 molecules has been identified⁴. It has been suggested that in CD8⁺ T cells the CD4 genes may be methylated and thus highly repressed⁵, whereas in CD4⁺ T cells the CD8 genes are unmethylated⁶ and their transcription can be induced by physiological stimuli such as interleukin-4 (ref. 7). Here, we demonstrate that infection with human herpesvirus 6 (HHV-6) (ref. 8), a virus proposed as a potential cofactor in AIDS (ref. 9), dramatically upregulates the expression of CD4—the receptor for human immunodeficiency virus type-1 (HIV-1) (refs 10–12)—in a human neoplastic T-cell line. More importantly, HHV-6 induces *de novo* expression of CD4 messenger RNA and protein in normal mature CD8⁺ T lymphocytes, rendering them susceptible to infection with HIV-1. These findings demonstrate that human CD3⁺CD4⁻CD8⁺ T lymphocytes can reacquire CD4 in the post-thymic life and elucidate a novel mechanism—receptor regulation—through which HHV-6 may positively interact with HIV-1 in coinfecting patients.

We initially studied the modulation of CD4 induced by HHV-6 in the human neoplastic T-cell line Jurkat, which has a characteristically low basal level of expression of CD4. During productive HHV-6 infection, a dramatic upregulation of CD4 was observed by indirect immunofluorescence analysis and fluorocytometry (Fig. 1), in parallel with the expression of HHV-6 antigens. Treatment of Jurkat cells with conditioned medium from human umbilical cord blood mononuclear cells stimulated with phytohemagglutinin (PHA) had no effect on CD4 expression (not shown), whereas the combined action of PHA and phorbol myristate-13-acetate (PMA) induced down-regulation of the antigen, as previously reported for PMA alone¹³. To evaluate whether expression of the immediate early genes of HHV-6 was sufficient to induce upregulation of CD4, we infected Jurkat cells in the continuous presence of 100 µg ml⁻¹ phosphonoformic acid (PFA), an inhibitor of the viral DNA polymerase that abrogates the synthesis of late herpesvirus gene products¹⁴. A consistent upregulation of CD4 was observed in PFA-treated cells, despite the complete lack of expression of the capsid (late) HHV-6 antigens, as recognized by monoclonal antibody 9A5D12 (ref. 15).

To investigate whether CD4 could be upregulated by other human herpesviruses, Jurkat cells were infected with either herpes simplex virus type-1 (HSV-1) or human cytomegalovirus

(hCMV). As previously demonstrated¹⁶, hCMV infection of T-lymphoid cells was non-productive, virus expression being restricted to the immediate early genes. By contrast, HSV-1 underwent a fully productive replication cycle, as revealed by both electron microscopy and induction of cytopathic effect. No alteration of CD4 expression occurred during infection with HSV-1 or hCMV (Fig. 1). Analogously, the CD4 antigen was not detected in the B-lymphoid cell line Raji after productive infection with a replication-competent Epstein-Barr virus (EBV) strain (not shown).

We next analysed mature CD4⁻CD8⁺ T lymphocytes, which lack CD4 mRNA and CD4 protein. Normal CD3⁺CD8⁺ T-cell populations were purified (<0.1% residual CD4⁺ cells) from adult peripheral blood by immunomagnetic beads. Productive infection of these cells with HHV-6 induced *de novo* expression of CD4, although expression of the CD8 and CD2 antigens was unaltered (Fig. 2a). At any time, the number of CD4⁺ cells in the cultures correlated well with the number of cells expressing HHV-6 antigens. For example, 8 days after infection, 51.0% of

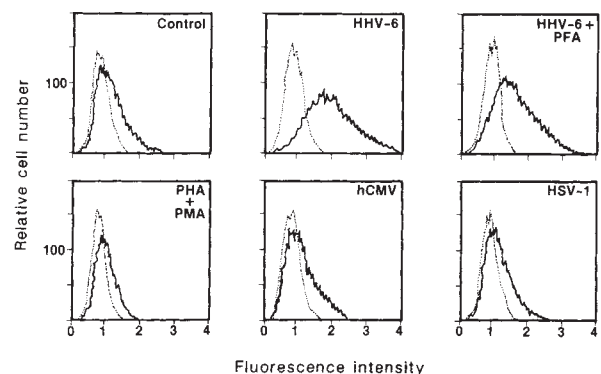
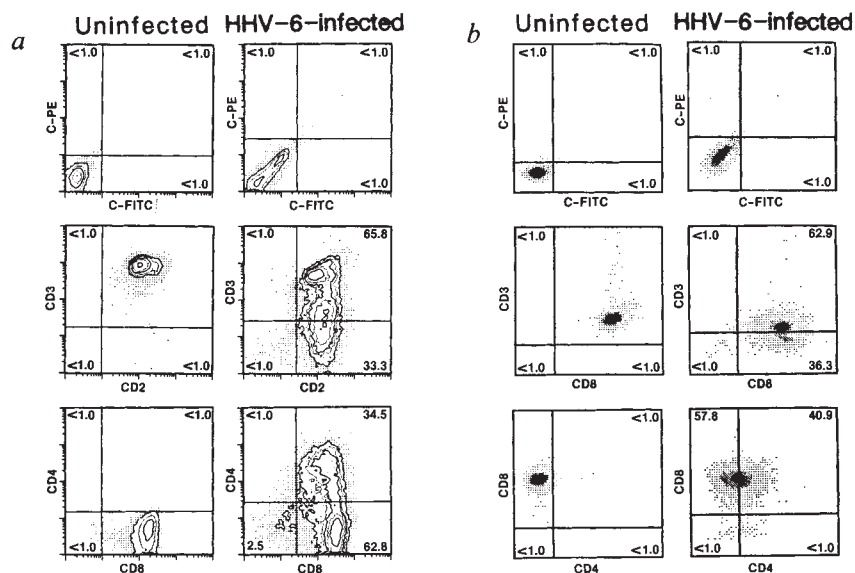


FIG. 1 Expression of CD4 by the human neoplastic T-cell line Jurkat after infection with the herpesviruses HHV-6, HSV-1 or hCMV, or stimulation for 24 h with PHA and PMA. Infection with HHV-6 was also performed in the continuous presence of 100 µg ml⁻¹ phosphonoformic acid (PFA), to restrict viral expression to the immediate early genes. Control uninfected Jurkat cells are represented in the upper left panel.

METHODS. The infection procedure employed was as described¹⁷. The cells were exposed to the supernatant of human umbilical cord blood mononuclear cells infected with HHV-6 (strain GS)⁸ for 1 h at 37 °C at a multiplicity of infection (MOI) of roughly 1. HSV-1 and hCMV were obtained from ATCC (Rockville, Maryland). The drug PFA was placed into the culture during the infection period and kept at the same original concentration (100 µg ml⁻¹) thereafter. Treatment with PFA alone did not alter CD4 expression. Controls included uninfected Jurkat cells washed and cultured in the same way as infected ones, and Jurkat cells costimulated with 2 µg ml⁻¹ PHA and 25 ng ml⁻¹ PMA for 24 h. Fluorocytometric analysis was performed at the time of maximal virus expression for each virus used, as determined by indirect immunofluorescence on acetone-fixed cells, performed as described¹⁷. For HHV-6, the cells were collected 6 days after infection (>70% HHV-6⁺); for HSV-1, 3 days after infection (56% HSV-1⁺); for hCMV, 10 days after infection (25% hCMV⁺). Infection by HHV-6 and HSV-1 was fully productive, whereas hCMV expression was restricted to immediate early genes. The analysis of the distribution of CD4 was performed on an Epic Profile (Coulter, Hialeah, Florida). The cells were stained with fluorescein-isothiocyanate (FITC)- or phycoerythrin (PE)-conjugated monoclonal antibodies. Controls (for each stimulus used) were stained with irrelevant mouse IgG₁ and IgG₂ conjugated to either of the fluorochromes. The results are expressed as arbitrarily normalized histograms (that is, number of cells against fluorescence intensity). At least 10,000 events were accumulated in all the tests. Expression of CD4 was monitored using four monoclonal antibodies (not all shown in the figure) binding two different domains of the CD4 glycoprotein^{23,24}. The antibodies used were for CD4 Leu3a and Leu3b (shown in the figure; Becton-Dickinson), OKT4a and OKT4 (Ortho Diagnostics, Raritan, New York), for HHV-6 9A5D12 (a kind gift of N. Balachandran)¹⁵ which recognizes a viral capsid component, for HSV-1 NEA-9251 (Dupont) (information on the antigen(s) recognized was not provided by the manufacturer) and for hCMV NEA-9221 (Dupont) which recognizes an early gene product.

FIG. 2 Fluorocytometric analysis of surface membrane antigen expression in normal human $CD3^+CD4^-CD8^+$ peripheral blood T lymphocytes after infection by HHV-6. Negatively selected $CD4^-CD8^+$ mononuclear cell population from normal adult peripheral blood (a) and $CD3^+CD4^-CD8^+$ T-cell clone number 20 (b) before and after HHV-6 infection. We tested purified cells from three different donors, containing more than 99.0% $CD3^+CD8^+$ T lymphocytes and less than 0.1% residual $CD4^+$ cells. The $CD8^+$ normal T-cell populations were obtained from adult human peripheral blood by negative selection after repeated rosetting of $CD4^-CD14^-CD19^-CD20^-CD56^-$ mononuclear cells with immunomagnetic beads (Dyna), as described previously¹⁷. The $CD8^+$ and $CD4^+$ T-cell clones were established by a high-efficiency limiting dilution from adult human peripheral blood, as previously described²⁵. The monoclonal antibodies used for fluorocytometric analysis were Leu2a (anti- $CD8$), Leu3a and Leu3b (anti- $CD4$), Leu4 (anti- $CD3$) and Leu5 (anti- $CD2$). OKT4 and OKT4a were also used in indirect immunofluorescence tests not shown in the figure. FITC- and PE-conjugated IgG₁ antibodies were used as controls (C-FITC and C-PE, respectively). The $CD4^-CD8^+$ mononuclear cell population was analysed with a FacScan analyser (Becton-Dickinson Immunocytometry, Mountain View, California), the T-cell clone with an Epic Profile. For details on fluorocytometric analysis, see the legend to Fig. 1.



the cells expressed viral antigens, as revealed by indirect immunofluorescence, whereas CD4 was detected in 34.5%. Two-colour fluorocytometry indicated that all the cells that had acquired CD4 also expressed CD8. As previously observed¹⁷, the CD3 antigen was significantly downregulated after HHV-6 infection, suggesting that the induction of CD4 was not the result of a nonspecific augmentation of gene expression.

To prove that the observed CD4 induction was not due to the selective outgrowth and differentiation of a minor subset of immature double-negative T cells or, alternatively, to the presence of residual $CD4^+$ cells, we studied T-lymphocytic clones of $CD3^+CD4^-CD8^+$ phenotype. These clones were obtained from normal adult peripheral blood by the limiting dilution technique. In all the clones tested ($n=5$), productive HHV-6 infection induced *de novo* expression of CD4 antigens. The fluorocytometric profile of CD3, CD4 and CD8 expression in a representative $CD8^+$ T-cell clone (number 20) is shown in Fig. 2b. In this clone, the percentage of HHV-6-expressing cells 8 days after infection was 58.0%, whereas CD4 was expressed on 40.9%. By contrast, the expression of CD8 was unaltered. Consistently, HHV-6 failed to induce *de novo* CD8 expression in any of seven $CD3^+CD4^-CD8^-$ T-cell clones tested as controls.

In further studies, HHV-6 also induced *de novo* CD4 expression in a human neoplastic T-cell line (HSB-2) displaying an immature $CD2^-CD3^-CD4^-CD7^+CD8^-$ phenotype, and in a $CD8^+$ mature T-cell clone (67-I) transformed with human T-cell lymphoma virus type-1 (not shown).

Inactivation of HHV-6 before infection, by either heat (56 °C for 40 min) or ultraviolet light (16 J m²), abrogated both the induction of CD4 in normal $CD4^-CD8^+$ T cells and the up-regulation of CD4 in Jurkat cells (not shown).

To elucidate whether induction of CD4 by HHV-6 occurs at the transcriptional level, we analysed northern blots of RNA extracted from normal $CD4^-CD8^+$ T-cell populations before and after infection with HHV-6. As expected, no signal was detectable in uninfected cells. By contrast, the 3.0-kilobase (kb) signal specific for CD4 mRNA was induced in $CD8^+$ T cells 8 days after infection (Fig. 3a). Simultaneous analysis of HHV-6 mRNA expression using a specific DNA probe (pZVH43) demonstrated that the induction of CD4 mRNA parallels the

transcription of viral genes (Fig. 3b). These findings suggest that the mechanism of CD4 induction by HHV-6 acts at the transcriptional level.

Because HHV-6 has been proposed as a potential cofactor in AIDS (ref. 9) and CD4 is the major membrane receptor for HIV-1 (refs 10–12), we investigated whether prior infection with HHV-6 enables HIV-1 to infect $CD4^-CD8^+$ T cells which alone are not susceptible to HIV-1. Normal $CD4^-CD8^+$ T lymphocytic populations infected 2 days earlier with HHV-6 were exposed for 2 h to HIV-1 (strain HTLV-III_B), then washed repeatedly and pelleted for DNA analysis by the polymerase chain reaction

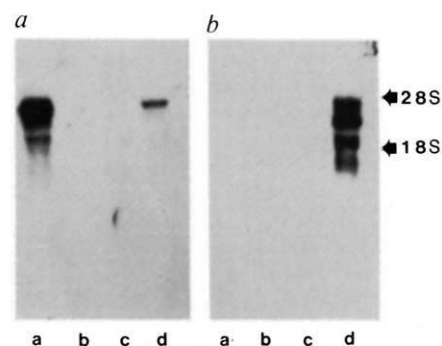
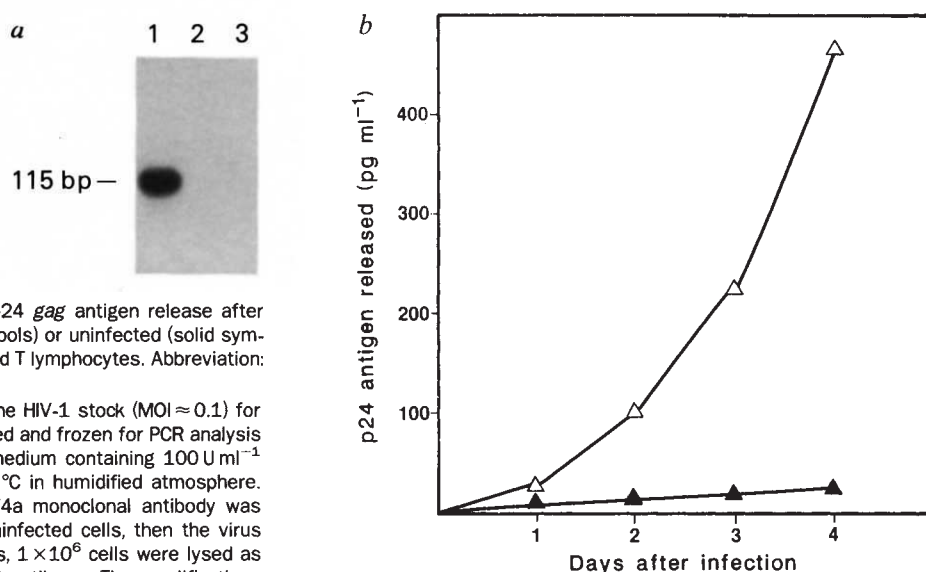


FIG. 3 Northern blot analysis of the expression of CD4 mRNA (a) and HHV-6 (b) before and after HHV-6 infection. Lanes a, control $CD4^+$ T-cell line CCRF-CEM; lanes b, control $CD4^+$ B-cell line Raji; lanes c, $CD3^+CD8^+$ normal human T lymphocytes; lanes d, HHV-6-infected $CD3^+CD8^+$ normal human T lymphocytes.

METHODS. Total cellular RNA was extracted from uninfected and HHV-6-infected cells (8 days after infection) as described²⁶ and 10 µg was separated by electrophoresis in 1.5% formaldehyde-agarose gel²⁷. After blotting on nylon membrane, RNA was cross linked by ultraviolet light treatment (1,200 µW cm²)²⁷ and then hybridized with a ³²P-labelled 1.8-kb *EcoRI*-*Bam*HI fragment of the CD4 cDNA pT4B (ref. 29) (a), or a ³²P-labelled 8-kb HHV-6-specific probe, pZVH43 obtained in our laboratory (b). A 3.0-kb band was expected for CD4, whereas four bands of approximate relative molecular mass 3.4, 2.7, 2.2 and 1.9 kb were expected for HHV-6. Positions of 18S and 28S ribosomal RNAs are shown.

FIG. 4 HIV-1 infection of normal human CD8⁺ T cells before and after induction of CD4 by HHV-6. *a*, Analysis by the polymerase chain reaction (PCR) to detect HIV-1 entry and reverse transcription in: HHV-6-infected normal CD3⁺CD4⁺CD8⁺ T lymphocytes (lane 1); HHV-6-infected normal CD3⁺CD4⁺CD8⁺ T lymphocytes plus OKT4a monoclonal antibody used at 25 $\mu\text{g ml}^{-1}$ (lane 2); uninfected normal CD3⁺CD4⁺CD8⁺ peripheral blood T lymphocytes (lane 3). *b*, Kinetics of HIV-1 p24 *gag* antigen release after HIV-1 infection of HHV-6-infected (open symbols) or uninfected (solid symbols) normal CD3⁺CD4⁺CD8⁺ peripheral blood T lymphocytes. Abbreviation: bp, base pairs.



METHODS. Cells (10^6) were incubated with the HIV-1 stock (MOI ≈ 0.1) for 2 h at 37 °C, then washed three times, pelleted and frozen for PCR analysis or, alternatively, resuspended in complete medium containing 100 U ml⁻¹ human delectinated IL-2 and cultured at 37 °C in humidified atmosphere. For the competition experiments, the OKT4a monoclonal antibody was preincubated for 30 min at 4 °C with the uninfected cells, then the virus was added for 2 h at 37 °C. For PCR analysis, 1×10^6 cells were lysed as described³⁰ and the lysate stored at -70 °C until use. The amplifications by PCR were carried out with: 100 μl reaction mixture, containing 50 μl cell lysate, 10 mM Tris buffer, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 200 μM each dNTP (dATP, dCTP, dGTP, dTTP), 10 pmol each primer and 2.5 U *Taq* polymerase (Beckman). The primers used were *gag* SK38 (ATAATCCACCTATCCAGTAGGAGAAAT) and *gag* SK39 (TTTGGTCCTGTCTTATGTCAGAAATG). Thirty cycles were run, including three steps of 2 min each at 94 °C, 55 °C and 72 °C, respectively. The extension step at 72 °C was prolonged 7 min after the last cycle. Each sample (20 μl) was loaded onto a 10% PAGE gel, run at 70 V for 3 h in TBE buffer and then electroblotted in the same buffer at 20 V overnight on a nylon membrane. The membrane

was prehybridized for 1 h at room temperature in 6 $\times$ SSC, 0.1% SDS, hybridized in the same buffer at room temperature overnight using ³²P-labelled SK19 *gag* as a probe (ATCCTGGGATTAATAAATAGTAAGAATGT-ATAGCCCTAC) and washed 2 $\times$ 20 min at room temperature in 6 $\times$ SSC and 2 $\times$ 20 min at 37 °C in the same buffer. The autoradiograms were exposed overnight at -70 °C. To monitor the HIV-1 p24 antigen release, we used an enzyme-linked immunosorbent assay (Organon Technika, Durham, North Carolina) with serially diluted antigen standards for quantitative measurement. The test was performed on cell-free supernatants from the cultures collected every day after HIV-1 infection.

or cultured in complete medium supplemented with 100 U ml⁻¹ of human interleukin-2 (IL-2). The same cells uninfected were simultaneously exposed to HIV-1 as a control. HIV-1 DNA was identified in HHV-6-infected CD8⁺ T cells, but not in control cells (Fig. 4a). To prove that CD4 is the receptor for HIV-1 in these HHV-6-infected cells, competition experiments were performed with the OKT4a monoclonal antibody. This antibody, used at 25 $\mu\text{g ml}^{-1}$, completely abrogated the penetration and reverse transcription of HIV-1 (Fig. 4a). To determine whether HIV-1 infection was productive, supernatants from infected cultures were monitored at various days after infection by an HIV-1 p24 antigen capture assay. The results (Fig. 4b) demonstrate that CD8⁺ T cells previously infected with HHV-6 can be productively infected with HIV-1.

Our findings indicate that mature CD4⁺CD8⁺ T lymphocytes can be induced by a naturally occurring agent to regain expression of CD4 in their post-thymic life. Although the mechanism of CD4 regulation by HHV-6 is presently unknown, it is probable that virus-specific or virus-induced factors activate

CD4 at the transcriptional level. Similar observations have been reported only for the use of stimuli not commonly encountered *in vivo*, such as the demethylating agent 5-azacytidine⁵ or the mitogen concanavalin A (ref. 18). Among the human *herpesviridae*, the ability to regulate CD4 seems to be restricted to HHV-6. Although human B cells transformed with EBV often exhibit CD4, induction of CD4 could not be demonstrated in normal cord blood B lymphocytes after exposure to EBV *in vitro*¹⁹. We have shown that by inducing CD4 and/or, possibly, other molecules involved in the HIV-1 receptor mechanism, HHV-6 expands the range of HIV-1 susceptible cells. In particular, HIV-1 can use the newly acquired CD4 to infect the CD8⁺ T lymphocyte. This may account for the defects observed in the CD8⁺ T-cell population in patients with advanced HIV-1 infection²⁰⁻²². But these defects could also derive from the abated 'helper' T-cell function, rather than from direct infection by cytopathic agents. Studies *in vivo* will be needed to elucidate the role of HHV-6 and its CD4-inducing ability in the natural history of HIV-1 infection. □

Received 1 July; accepted 12 December 1990.

- Reinherz, E. L. & Schlossman, S. F. *Cell* **19**, 821-827 (1980).
- Scollay, R., Bartlett, P. & Shortman, K. *Immun. Rev.* **89**, 79-95 (1984).
- von Boehmer, H. et al. *A. Rev. Immun.* **6**, 309-328 (1988).
- Blue, M. L., Daley, J. F., Levine, H. & Schlossman, S. F. *J. Immun.* **134**, 2281-2286 (1985).
- Richardson, B., Kahn, L., Lovett, E. J. & Hudson, J. *J. Immun.* **137**, 35-39 (1986).
- Carbone, A. M., Marrack, P. & Kappler, J. W. *Science* **242**, 1174-1176 (1988).
- Pallard, X., de Waal Malefijt, R., de Vries, J. E. & Spits, H. *Nature* **335**, 642-644 (1988).
- Salahuddin, S. Z. et al. *Science* **234**, 596-601 (1986).
- Lusso, P. et al. *Nature* **337**, 370-373 (1989).
- Dalgleish, A. G. et al. *Nature* **312**, 763-767 (1984).
- Klatzman, D. et al. *Nature* **312**, 767-768 (1984).
- McDougal, J. S. et al. *Science* **231**, 382-385 (1986).
- Solbach, W. *J. exp. Med.* **156**, 1250-1263 (1982).
- Stinski, M. F. *J. Virol.* **23**, 751-767 (1977).
- Balachandran, N., Amelse, R. E., Zhou, W. W. & Chang, C. K. *J. Virol.* **63**, 2835-2840 (1989).
- Rice, G. P. A., Schrier, R. D. & Oldstone, M. B. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6134-6138 (1984).

- Lusso, P. et al. *J. exp. Med.* **167**, 1659-1670 (1988).
- Blue, M. L., Daley, J. F., Levine, H., Craig, K. A. & Schlossman, S. F. *J. Immun.* **137**, 1202-1207 (1986).
- Pagano, J. S., Kenney, S., Markovitz, D. & Kamine, J. *J. Virol. Meth.* **21**, 229-239 (1988).
- Shearer, G. M. et al. *J. clin. Invest.* **76**, 1699-1704 (1985).
- Gerstoft, J., Dickmeiss, E. & Methiesen, L. *Scand. J. Immun.* **22**, 463-469 (1985).
- Sharma, B. & Gupta, S. *Clin. exp. Immun.* **62**, 296-303 (1985).
- Bach, M.-A. et al. *J. Immun.* **132**, 735-740 (1984).
- Sattentau, Q. J., Dalgleish, A. G., Weiss, R. A. & Beverley, P. C. L. *Science* **234**, 1120-1123 (1986).
- Moretta, A., Pantaleo, G., Moretta, L., Cerottini, J. C. & Mingari, M. C. *J. exp. Med.* **157**, 743-751 (1983).
- Wilkinson, M. *Nucleic Acid Res.* **16**, 10933 (1988).
- Biassoni, R., Ferrini, S., Prigione, I., Moretta, A. & Long, E. O. *J. Immun.* **140**, 1685-1690 (1988).
- Feinberg, A. P. & Vogelstein, S. *Ann. Biochem.* **132**, 6-12 (1983).
- Maddon, P. J. et al. *Cell* **42**, 93-104 (1985).
- Psallidopoulos, H. C. et al. *J. Virol.* **63**, 4626-4633 (1989).

ACKNOWLEDGEMENTS. We are thankful to A. S. Fauci and E. O. Long for reviewing the manuscript.

Evidence from reversal of handedness in *C. elegans* embryos for early cell interactions determining cell fates

W. B. Wood*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK, and Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80302, USA

MANY animals with overall bilateral symmetry also exhibit some left–right asymmetries with generally invariant handedness. Therefore, the left–right embryonic axis must have a consistent polarity, whose origins and subsequent effects on development are not understood (reviewed in ref. 1). *Caenorhabditis elegans* exhibits such left–right asymmetries at all developmental stages. The embryonic cell lineage is asymmetric as well: although the animal is generally bilaterally symmetric, many of its contralaterally analogous cells arise from different lineages on the two sides of the embryo^{2,3}. I accomplished reversal of embryonic handedness by micromanipulation at the 6-cell stage, which resulted in mirror-image but otherwise normal development into healthy, fertile animals with all the usual left–right asymmetries reversed. This result demonstrates that in the 6-cell embryo the pair of anterior (AB) blastomeres on the right is equivalent to the pair on the left, and that the extensive differences in fates between lineally homologous derivatives of these cells on the two sides of the animal must be dictated by cell interactions, most of which are likely to occur early in embryogenesis.

The body plan of the *C. elegans* adult hermaphrodite (Fig. 1) is typical for nematodes. There are a few left–right (l–r) asymmetries, but the majority of tissues and cells are arranged with bilateral symmetry^{4–6}. The embryo, by contrast, shows marked l–r asymmetry, becoming only gradually more symmetric as embryogenesis progresses. Handedness seems to be invariant: *C. elegans* embryos with reversed handedness have never been reported.

Bilateral asymmetry first becomes evident between the 4- and 6-cell stages. Left–right polarity cannot be fixed until after dorsal–ventral polarity is established between the 2- and 3-cell stages⁷, because switching the positions of the two AB cells in the 3-cell embryo by micromanipulation reversed the dorsal–ventral axis but did not lead to handedness reversal. The 4-cell embryo is planar (Fig. 2A, B) and apparently bilaterally symmetric. In the next round of cell division, skewing of the l–r cleavages of ABa and ABp results in positioning of the al and pl daughters somewhat anterior to ar and pr, respectively (Figs 2C, 3d). From this stage onward, there are substantial differences between the two sides of the embryo, not only in the positions of lineal homologues but also in the cell lineages by which they produce progeny cells of similar fates on either side of the animal². This is particularly true of the ABa daughters, which give rise to much of the anterior nervous system as well as muscles of the pharynx: ABal and ABar, although generating similar sets of contralateral analogues, do so by quite different lineal patterns^{2,3}.

This piecemeal generation of symmetry by an asymmetric lineage³ could be accomplished in either of two ways: by cell-autonomous (lineal) programming that has simply evolved differently for AB descendants on the two sides of the animal, or alternatively by cell interactions that dictate similar cell fates

in contralaterally equivalent positions. These alternatives should be distinguishable experimentally if the relative anterior–posterior positions of the cell pairs ABal-pl and ABar-pr at the 6-cell stage (Fig. 2C) could be reversed, by moving ABal-pl back and ABar-pr forward, thereby changing the handedness of the embryo. If the l–r differences in cell fates are lineally programmed, then grossly abnormal development should result as the left and right lineages produce their respective patterns of cell fates in the wrong relative anterior–posterior positions. But if l–r differences are dictated by cell interactions, then a mirror-image but otherwise normal lineage might be generated.

Handedness reversal proved possible to accomplish by micromanipulation during cleavage of ABa and ABp as shown in Fig. 3a and b. This procedure reversed the normal skewing of both AB spindles and resulted in ar and pr being anterior to al and pl, respectively, at completion of the cleavage (Fig. 3c).

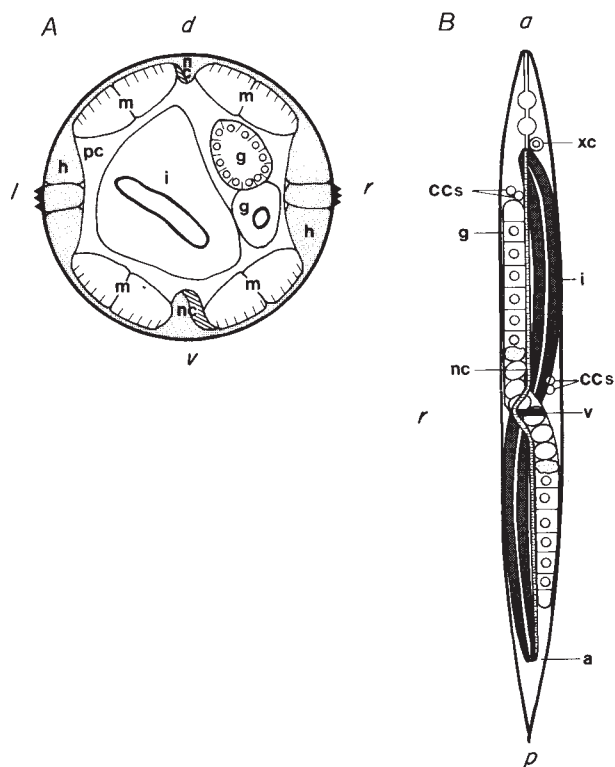


FIG. 1 Bilateral symmetry and l–r asymmetries in the adult *C. elegans* hermaphrodite. **A**, Cross-section through anterior, between pharynx and uterus, viewed from the rear. **B**, Ventral view of interior organs. Anterior, posterior, dorsal, ventral, left and right are indicated by a, p, d, v, l and r, respectively. The bilobed gonad (g) and the intestine (i) lie in the pseudocoelom (pc) with the anterior gonad lobe to the right and the posterior lobe to the left of the intestine (**A, B**), which terminates at the anus (a). The excretory cell nucleus (xc) lies to the left of the midline; the anterior pair of coelomocytes (CCs) is located on the right and the posterior pair on the left of the pseudocoelomic cavity (**B**). The ventral nerve cord (nc) lies on the midline except that it passes around the vulva (v) on the right side (**B**). In the ventral nc the cell bodies all lie to the right of the neuronal processes, which run along the left side of the cord; the dorsal nc shows the opposite asymmetry (**A**)⁵. The gonad and intestine asymmetries can be clearly seen with a dissecting microscope if animals are rolled 90° from their normal orientation (on one side or the other) to ventral- or dorsal-side up (for example, Fig. 3e); the remaining asymmetries were observed under a Zeiss compound microscope equipped with Nomarski optics, in animals mounted¹⁷ ventral-side up. No animals with reversed gonad handedness were seen in examination of more than 2,500 wild-type hermaphrodites reared at 16°. Two hermaphrodites with several reversed asymmetries have been observed, however, by J. Sulston (personal communication) in the course of examining a large number of fixed and stained animals. Animals with reversed asymmetries also have been obtained by the author from embryos treated with chitinase to remove the egg shell and among the progeny of mutagenized populations (unpublished observations).

* Present address: Department of Molecular, Cellular, and Developmental Biology, Box 347, University of Colorado, Boulder, Colorado 80309, USA.

The subsequent divisions of the EMS and P_2 cells gave a mirror-image 8-cell embryo (not shown) in which the normal asymmetric positions of non-AB cells (see Fig. 2D) were also reversed, with MS to the left and C to the right of the midline. Reversed embryos underwent subsequent cleavages on schedule and hatched as viable larvae that grew into reversed but fertile adults (Fig. 3e).

The embryonic development of a reversed embryo was observed and recorded for subsequent lineaging using a programmable video disc system (J. White, personal communication). Timing of divisions and positioning of specific cells in the reversed embryo were consistent with an enantiomorphic but otherwise normal pattern of development, as illustrated by examples from two stages in Fig. 4. In the first-stage (L1) larva that hatched from this embryo, 14 normally asymmetrically placed nuclei⁴ were scored, representing descendants of founder cells AB, MS, C and P_4 (see legend to Fig. 4). All were found in their usual anterior-posterior positions, but on the opposite side of the animal than normally (not shown). The adults that developed from reversed embryos exhibited reversed handedness for all the cells and tissues shown in Fig. 1. These animals were healthy, moved normally and produced normal numbers of self progeny, all of which showed normal l-r asymmetry.

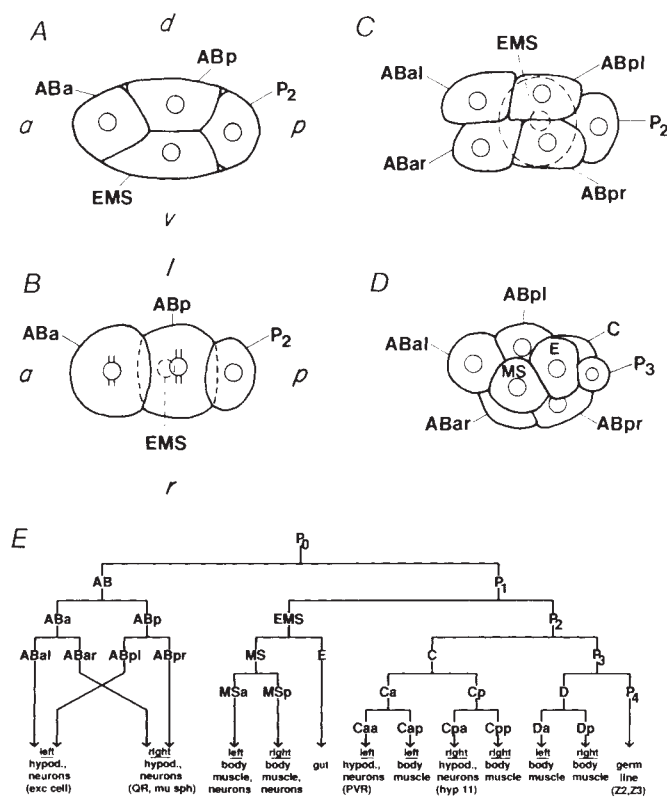


FIG. 2 Cell orientations in 4-cell to 8-cell embryos. A, Four-cell embryo, left lateral view (see E for explanation of cell nomenclature). B, Four-cell embryo, ventral view focused on lower plane. Spindles in ABa and ABp cells are forming perpendicular to the a-p axis and will become skewed in an anticlockwise direction as cleavage proceeds. Dashed lines show positions of EMS cell and its nucleus in upper focal plane. C, Six-cell embryo, ventral view focused on lower plane. AB cells are arranged asymmetrically as a result of the normal skewing, with ABal and pl daughters somewhat anterior to ABar and pr, respectively. ABal is also somewhat ventral to ABar. D, Eight-cell embryo, ventral view. EMS and P_2 have divided, also asymmetrically, so that MS and P_3 are to the right and E and C to the left of the midline. E, Lineage diagram of early cleavages showing origins of the six founder cells AB, MS, E, C, D and P_4 and major developmental fates of their progeny. Relative timing of divisions on the vertical axis are not accurately depicted. Progeny of founder cells are named according to their positions following cleavage; for example, ABa is the anterior daughter of AB; ABal is the left daughter of ABa, and so on.

Three significant conclusions can be drawn from these results. (1) Handedness of the embryo at the 6-cell stage determines handedness of subsequent asymmetries throughout development and in the adult. (2) At this stage, ABal must be developmentally equivalent to ABar and ABpl to ABpr; the differences in the fates of these pairs of cells cannot be attributed to intrinsic 'left' and 'right' determinants segregated to the appropriate daughters in the ABa and ABp cleavages. (3) Therefore, the extensive differences in fates of lineal homologues on the two sides of the embryo^{2,3} must be determined by cell interactions, which differ on the left and right because of the asymmetric positioning of AB derivatives relative to each other and to other cells in the embryo.

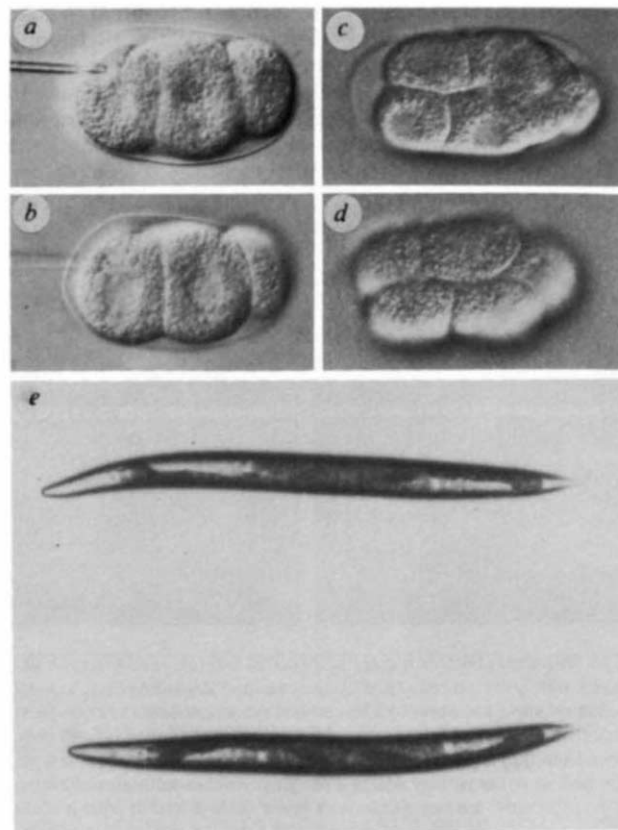


FIG. 3 Handedness reversal: operation and results in embryos and adults. a, Nomarski photomicrograph of 4-cell embryo, ventral side up as in Fig. 2B, with microneedle in position for operation (see below). b, Same embryo as in a, lower focal plane showing ABA nucleus at onset of spindle formation. c, Reversed 6-cell embryo following successful operation, focused near lower surface to show cleavage planes optimally. d, Normal 6-cell embryo, similar stage and focal plane to c. e, The reversed (above) and normal (below) adult hermaphrodites that developed from the operated and control embryos shown in c and d, respectively, viewed ventral-side up. Embryonic handedness reversal, judged by configuration at the 6-cell stage (c, d), was achieved with a success rate of about 25%, probably depending most critically on timing and orientation of the embryo. All reversed embryos (6/6) developed into reversed but otherwise normal and fertile adults. About a dozen unsuccessfully operated embryos, which appeared as in d at the 6-cell stage, were incubated as controls; all developed into fertile adults with normal handedness.

METHODS. *C. elegans* was cultivated and embryos obtained by standard procedures¹⁷. Microneedles were prepared and embryos mounted for micromanipulation according to ref. 7. Embryos were rolled to ventral-side-up at the 4-cell stage, and then downward and rearward pressure was exerted on the left ventral surface of the ABa cell as in a, during spindle formation and subsequent cleavage of ABa and ABp. The two adult hermaphrodites (e) were lightly anaesthetized in 0.7% phenoxypropanol¹⁷, rolled to ventral-side up on an agar surface, and photographed together using a Wild M400 microscope.

These interactions probably take place early in embryogenesis. The contacts of AB homologues at the 6-cell stage are still equivalent (Fig. 2C), but by the 8-cell stage they are different on the two sides of the embryo (Fig. 2D). In later embryos (>51 cells), laser ablation experiments² provided convincing evidence for cell autonomy of many AB-cell fates. Therefore, most of the determinative interactions are likely to occur between the 8-cell and 51-cell stages.

Conclusion (1) confirms a conjecture put forward nearly a century ago by zur Strassen⁸, who observed that about 2.5% of embryos from the parasitic nematode *Ascaris* normally have reversed l-r asymmetry and postulated that these would give rise to reversed adults. Conclusions (2) and (3) argue against the predominant classical view (see, for example, ref. 9; not shared by all early investigators, such as ref. 10) that nematode embryos are strictly mosaic in their mode of early cell determination. My experiments support the alternative view that *C. elegans* embryos are much more similar to those of vertebrates than previously assumed, in that only some aspects of cell determina-

tion can be attributed to autonomous determinants^{2,11-13}, whereas most others are dictated by intercellular signalling. Earlier evidence on this point came from Schierenberg¹⁴ and in particular Priess and Thomson⁷, who demonstrated equivalence of ABa and ABp at the 4-cell stage and implicated interactions of AB descendants with EMS descendants in the induction of AB-derived pharyngeal muscle cells. Also consistent with this view was the finding¹⁵ that mutations in the *glp-1* gene, which encodes a protein similar to known intercellular signalling components¹⁶, can block induction of AB-derived pharyngeal muscles. The results presented here indicate that extensive cell interactions are required in the early embryo for determination of the many lineal homologues that show l-r differences in developmental fate. □

Received 29 October; accepted 7 December 1990.

1. Brown, N. A. & Wolpert, L. *Development* **109**, 1-9 (1990).
2. Sulston, J. E. *et al.*, *Dev. Biol.* **100**, 64-119 (1983).
3. Sulston, J. E. *Cold Spring Harb. Symp. quant. Biol.* **48**, 443-452 (1983).
4. Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110-156 (1977).
5. White, J. G. *et al.*, *Phil. Trans. R. Soc.* **314**, 1-340 (1986).
6. White, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 81-122 (Cold Spring Harbor, Laboratory, New York, 1988).
7. Priess, J. & Thomson, J. N. *Cell* **48**, 241-250 (1987).
8. zur Strassen, O. *Arch. EntwMech. Org.* **3**, 27-105 (1896).
9. Dünschen, F. *Arch. EntwMech. Org.* **115**, 237-335 (1929).
10. Bonfig, R. Z. *Wissenschaft. Zool.* **124**, 407-456 (1925).
11. Laufer, J. S., Bazzicalupo, P. & Wood, W. B. *Cell* **19**, 569-577 (1980).
12. Cowan, A. E. & McIntosh, J. R. *Cell* **41**, 923-932 (1985).
13. Edgar, L. G. & McGhee, J. D. *Cell* **53**, 589-599 (1988).
14. Schierenberg, E. *Dev. Biol.* **122**, 452-463 (1987).
15. Priess, J. R., Schnabel, H. & Schnabel, R. *Cell* **51**, 601-611 (1987).
16. Yochem, J. & Greenwald, I. *Cell* **58**, 553-563 (1989).
17. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 587-606 (Cold Spring Harbor Laboratory, New York, 1988).
18. Deppe, U. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **75**, 376-380 (1978).

ACKNOWLEDGEMENTS. I am grateful for discussion, helpful suggestions and hospitality to J. White, J. Sulston, J. Rothman, A. Chisom and others in the Cell Biology Unit of the MRC Laboratory of Molecular Biology, where most of this work was done during a sabbatical leave from University of Colorado, Boulder. I also thank J. White for use of the 4-D video recording system, J. Rothman for optical disc recordings of normal embryonic development, and J. Rothman, J. White and C. Hunter for comments on the manuscript. This research was supported by NIH and MRC.

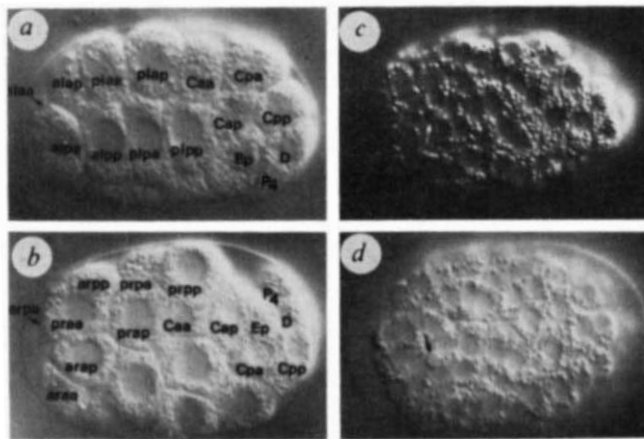


FIG. 4 Nomarski photomicrographs of normal and reversed embryos at two stages, with selected cells labelled. *a, b*, Normal 28-cell embryo, superficial left lateral view, and reversed 28-cell embryo, superficial right lateral view, respectively. Labelled cells with no founder-cell designation are AB descendants. Note that C-cell progeny, normally on the left side as in *a*, are on the right side in *b*; conversely, MS cell progeny, which would normally be seen on the right side, are not visible (in a lower focal plane) in both *a* and *b*. *c, d*, Superficial dorsal views of normal and reversed embryos, respectively, at about the 360-cell stage. White arrow in *c* points to ABapaaapp undergoing programmed cell death (reproduced, with permission, from ref. 2, Fig. 2d); the analogous dying cell in *d* was shown by lineaging to be ABalpaapp (black arrow), which in a normal embryo would give rise to a pharyngeal neuron². Nuclei scored for position in the L1 that hatched from this embryo were as follows. AB descendants: QR (previously Q2; ref. 4), the excretory cell, and the rectal sphincter muscle cell mu sph; MS descendants: Z1 and Z4 in the gonad primordium, the mesoblast M, and the four coelomocytes; C descendants: hyp11 and PVR in the tail; P₄ descendants: Z2 and Z3 in the gonad primordium. The first cleavages of MS and C are predominantly a-p in orientation¹⁸, although they also have a l-r component (J. Rothman, personal communication). For both MS and C, the *a* daughters probably give rise to the same cells in reversed embryos (not all lineages were followed) as in normal embryos, though in mirror-image l-r positions (for example, MSa in reversed embryos gives rise to right body muscle and neurons, and also to Z4 and M; compare Fig. 2E). The same is true for the *p* daughters of MS and C. Therefore, this experiment cannot distinguish between lineal programming or cell interactions in determining the l-r differences in these lineages.

METHODS. Embryos were mounted as described² and their development was recorded using the 'four-dimension' video recording system (see text), which consists of a Nomarski microscope (Zeiss Axioplan) equipped with a video camera and connected to a computer-controlled optical disc recorder (Sony) and focusing drive motor (Biorad). The system automatically records complete sets of serial optical sections at predetermined intervals and allows subsequent playback of consecutive images from any desired section to facilitate lineaging. Photographs were taken from the video monitor.

Requirement for the replication protein SSB in human DNA excision repair

Dawn Coverley, Mark K. Kenny, Maureen Munn*, W. Dean Rupp*, David P. Lane & Richard D. Wood

Imperial Cancer Research Fund, Clare Hall Laboratories, Blanche Lane, South Mimms, Hertfordshire EN6 3LD, UK

* Radiobiology Laboratories, Yale University School of Medicine, PO Box 3333, New Haven, Connecticut 06510-8039, USA

REPLICATION and repair are essential processes that maintain the continuity of the genetic material. Dissection of simian virus 40 (SV40) DNA replication has resulted in the identification of many eukaryotic replication proteins, but the biochemistry of the multienzyme process of DNA excision repair is less well defined. One protein that is absolutely required for semiconservative replication of SV40 DNA *in vitro* is human single-stranded DNA-binding protein (SSB, also called RF-A and RP-A)¹⁻³. SSB consists of three polypeptides of relative molecular mass 70,000, 34,000 and 13,000, and acts with T antigen and topoisomerases to unwind DNA, allowing the access of other replication proteins. Human SSB can also stimulate the activity of polymerases α and δ , suggesting a further role in elongation during DNA replication⁴⁻⁶. We have now found a role for human SSB in DNA excision repair using a cell-free system that can carry out nucleotide excision repair *in vitro*⁷. Monoclonal antibodies against human SSB caused extensive inhibition of DNA repair in plasmid molecules damaged

by ultraviolet light or acetylaminofluorene. Addition of purified SSB reversed this inhibition and further stimulated repair synthesis by increasing the number of repair events. These results show that a mammalian DNA replication protein is also essential for repair.

The extent of nucleotide excision repair can be assessed *in vitro* by incubating damaged and nondamaged circular plasmid DNA with whole-cell extracts⁸ from human cell lines in a reaction mixture which includes the four deoxynucleoside triphosphates and [α -³²P]dATP. Enzymes in the extract carry out all essential stages of excision repair: recognition of damage, cleavage of the DNA by incision proteins, excision of damaged

nucleotides, resynthesis and ligation⁷. Quantification of incorporated radioactive material yields a measure of DNA repair. Using this approach it has been shown that cell extracts can carry out repair synthesis in DNA damaged by ultraviolet light, psoralens and platinating agents^{7,9-11} and that this repair synthesis is localized to sites of DNA damage¹². Extracts from cell lines derived from patients with the inherited DNA excision repair disorder xeroderma pigmentosum (XP)¹³ are deficient in repair synthesis in this assay^{7,14}.

Four monoclonal antibodies directed against human SSB⁶ and known to inhibit SV40 DNA replication *in vitro* were tested for their effect on repair synthesis in plasmid DNA damaged by ultraviolet light. Three antibodies (70A, 70B and 70C) recognize the SSB subunit of relative molecular mass 70,000 (M_r 70K), and one (34A) recognizes the 34K subunit. All four antibodies inhibited repair synthesis mediated by HeLa cell extracts (Fig. 1) or by extracts from the normal lymphoid cell line GM1953 (not shown). Anti-SSB 70C and 34A antibodies were the most inhibitory, causing a reduction in DNA synthesis

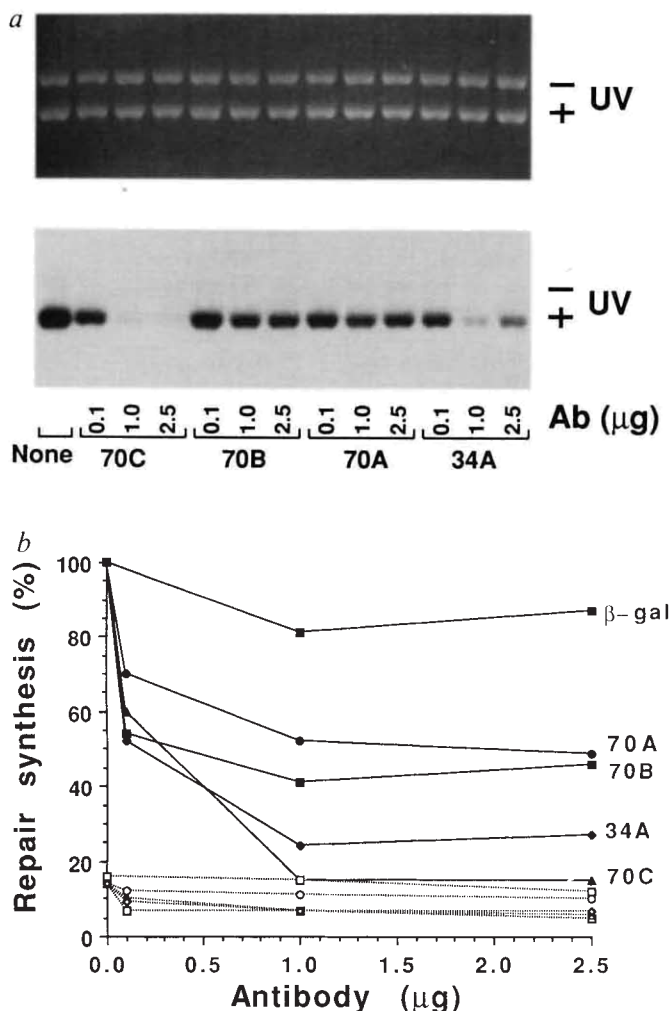


FIG. 1 Inhibition of DNA excision repair by anti-SSB monoclonal antibodies. *a*, DNA isolated from repair reaction mixtures was visualized by ethidium-bromide staining (top panel). Incorporation of radiolabelled nucleotides into this DNA was visualized by fluorography (lower panel). In the control reaction without antibody (Ab; lane 1) 164 fmol dAMP (100% synthesis) was incorporated into UV-irradiated DNA (lower band), and 23 fmol into unirradiated control DNA (upper band). *b*, Results were quantified as described⁷. Open symbols, unirradiated plasmid DNA; closed symbols, irradiated DNA. 70A, ●, ○; 70B, ■, □; 70C, ▲, △; 34A, ◆, ◇; control β-gal antibody, ■, □. For repair assays, 100 μg protein extracted from HeLa cells⁸ was preincubated with 0.0, 0.1, 1.0 or 2.5 μg of anti-SSB 70A, 70B, 70C or 34A purified monoclonal antibody⁶ for 30 min at 37 °C. UV-irradiated (450 J/m²) pAT153 plasmid DNA (3.7 kbp; 300 ng) and 300 ng non-irradiated pBR322 plasmid DNA (4.4 kbp) were then added to give a total volume of 50 μl in reaction buffer (45 mM HEPES-KOH (pH 7.8), 60 mM KCl, 7.5 mM MgCl₂, 0.9 mM dithiothreitol, 0.4 mM EDTA, 2 mM ATP, 20 μM each of dGTP, dCTP and TTP, 8 μM dATP, 2 μCi[α -³²P]dATP (3,000 Ci mmol⁻¹), 40 mM phosphocreatine, 2.5 μg creatine phosphokinase (type I, Sigma), 3.4% glycerol and 18 μg BSA). Reactions were allowed to proceed for 3 h at 30 °C. DNA was isolated, linearized by digestion with *Eco*RI, and separated by agarose gel electrophoresis⁷.

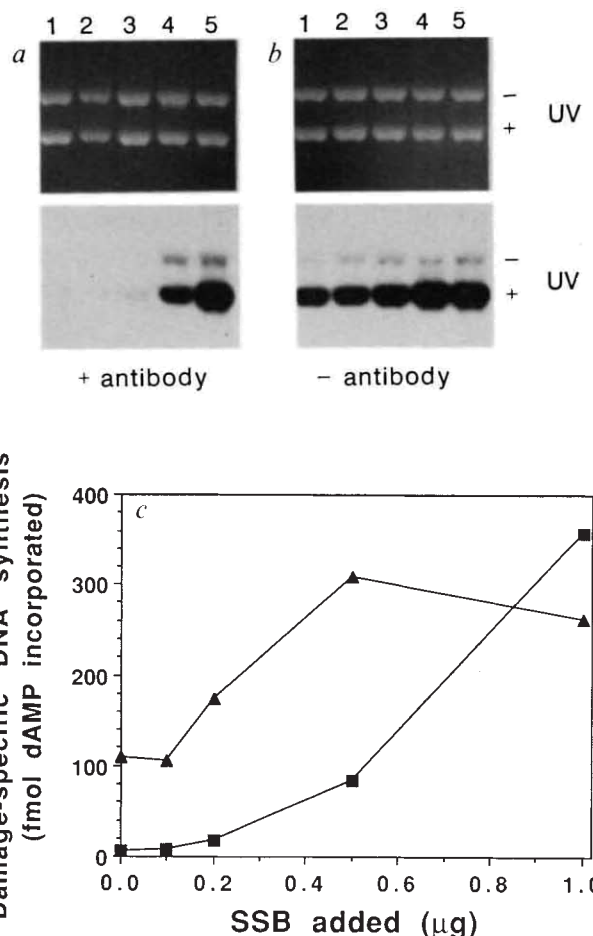


FIG. 2 Reconstitution of repair activity by the addition of purified human SSB. Extract protein (100 μg) from the lymphoblastoid cell line GM1953 (derived from a normal individual) was preincubated for 30 min at 37 °C with 0.0, 0.1, 0.2, 0.5 or 1.0 μg human SSB (400 U mg⁻¹; lanes 1–5) purified from HeLa cells⁶, and 1 μg anti-SSB 70C antibody (*a*) or no antibody (*b*). Ultraviolet-irradiated plasmid DNA and unirradiated control DNA were then added with the other components of the repair reaction mixture, and incubation proceeded for 3 h at 30 °C. *a*, *b*, Top panel, DNA stained with ethidium bromide. Bottom panel, incorporation of radiolabelled dAMP into this DNA as visualized by fluorography. *c*, Results were plotted as damage-specific synthesis (incorporation into damaged DNA minus incorporation into undamaged control DNA)⁷ in the presence (■) or absence (▲) of antibody. Similar reconstitution of repair activity by SSB was also achieved with antibody-inhibited HeLa cell extract.

to background levels. These results are comparable to those seen in the *in vitro* SV40 DNA replication system⁶.

The inhibition of DNA repair synthesis by anti-SSB antibodies was caused specifically by the neutralization of human SSB. Cell extract, in the presence or absence of anti-SSB antibody, was preincubated with varying amounts of pure human SSB⁶ and then assayed for repair activity. Human SSB was able to reverse the inhibitory effect of the 70C antibody (Fig. 2a) as well as the 70A, 70B and 34A antibodies (data not shown). Addition of more SSB than that required to bind all of the antibody led to further stimulation of repair synthesis (Fig. 2a,c). Moreover, supplementation of cell extract with human SSB in the absence of antibody increased ultraviolet damage-dependent DNA repair synthesis by about three times (Fig. 2b,c). Addition of more than 1 µg of SSB to the standard 50-µl reaction mixture did not further stimulate synthesis. Neither *Escherichia coli* SSB (Pharmacia) nor adenovirus DNA binding protein¹⁵ could reverse the inhibitory effect of anti-SSB antibodies (not shown).

To investigate further the role for SSB in DNA excision repair, human SSB was added to repair-defective extracts of cell lines derived from individuals with XP¹⁶. SSB did not affect the low level of repair synthesis in damaged DNA when added to extracts from cell lines of the XP complementation groups A, B, C, D, F and G (data not shown). An alteration in SSB is not responsible for the repair defect in groups A or B, because the XP-A and XP-B correcting genes encode distinct proteins of M_r 31K and 89, respectively^{17,18}. Our results suggest that an alteration in SSB also does not account for the repair defect in groups C, D, F or G. Further, the lack of effect of excess SSB on XP extracts may indicate that SSB is already present in

sufficient amounts to support the limited amount of repair that occurs in these extracts.

SSB may act in repair coordinately with the DNA repair polymerase(s). In mammalian cells, DNA polymerase ϵ (previously known as PCNA-independent DNA polymerase δ) has been implicated as the principal polymerase in nucleotide excision repair¹⁹. Human SSB can stimulate the activity of mammalian DNA polymerases^{4,5}, including polymerase ϵ (Z.-Q. Pan, S.-H. Lee and J. Hurwitz, personal communication). Thus, SSB might effect an increase in damage-specific DNA synthesis by increasing the size of repair patches. This could result from a direct stimulation of polymerase processivity, or by the promotion of strand displacement or exonucleolytic degradation at incised sites. The results with XP extracts argue against such an increase in repair patch size, as this should increase the residual synthesis seen with XP extracts by the same factor as with repair-proficient extracts. Alternatively, SSB could stimulate DNA repair synthesis by increasing the number of incision events.

To distinguish between these possibilities, we analysed the extent of repair synthesis at a single site. A double-stranded bacteriophage M13mp18 DNA molecule was used that contained one 2-(acetylaminofluorene) (AAF) adduct per circle at a defined G residue¹². Analysis of DNA repair synthesis in different restriction fragments of this molecule (Fig. 3a) reveals the pattern and extent of repair events. The majority (74%) of AAF adduct-dependent repair synthesis carried out by HeLa cell extract was confined to the 31-base pair (bp) *SphI*-*Bam*HI restriction fragment (Fig. 3b, lane 1). The single AAF adduct is located in the centre of this fragment. Additional damage-

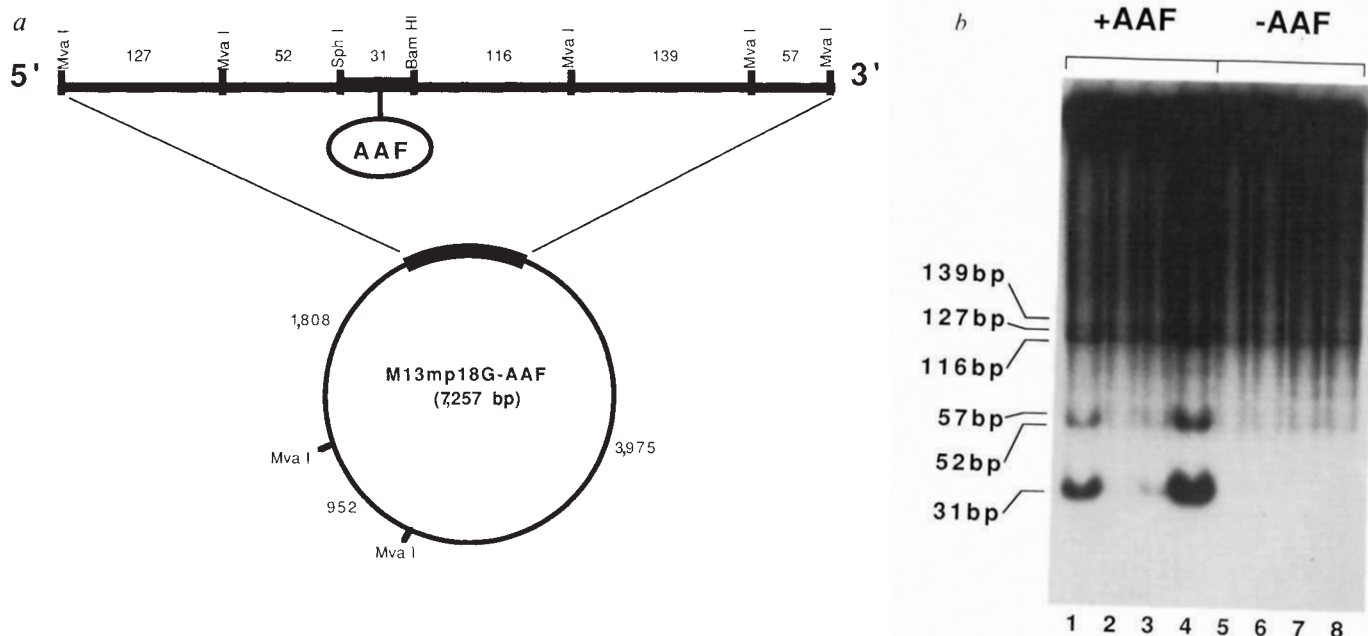


FIG. 3 Structure of repair patches in the presence and absence of SSB. a, The analysis makes use of closed circular double-stranded M13mp18G-AAF DNA, which contains an adduct of 2-(acetylaminofluorene) (AAF) at a single G residue in the centre of a 31-bp *SphI*-*Bam*HI restriction fragment¹². DNA constructed identically, but without the AAF adduct, was used as a control. HeLa whole-cell extract (100 µg protein) was pre-incubated for 30 min at 37 °C with anti-SSB 70C antibody and/or human SSB, as indicated. M13mp18G-AAF DNA (300 ng) and reaction buffer were then added. After 3 h at 30 °C, the DNA was recovered and digested with *SphI*, *Bam*HI and *MvaI* restriction endonucleases before electrophoresis on a 15% polyacrylamide gel. b, An autoradiograph of the dried gel. Lanes 1-4, reactions with M13mp18G-AAF DNA; lanes 5-8, reactions with M13mp18G DNA (control DNA which does not contain an AAF adduct). Lanes 1 and 5, HeLa extract only. Lanes 2 and 6, HeLa extract pre-incubated with 1 µg 70C antibody.

Lanes 3 and 7, HeLa extract pre-incubated with 1 µg 70C antibody and 0.5 µg human SSB. Lanes 4 and 8, HeLa extract pre-incubated with 1 µg human SSB in the absence of antibody. Repair synthesis in individual restriction fragments was assessed by densitometry of the autoradiograph using an LKB 2222 UltraScan XL laser densitometer and LKB GelScan XL version 2.0 software. The densitometric results were confirmed and calibrated by direct counting of β emissions from the dried gel, using a two-dimensional array of proportional counters (AMBIS Radioanalytic Imaging System) and associated software (AMBIS Systems, Inc., San Diego, California). All restriction fragments (from plasmids with or without AAF) show background DNA synthesis proportional to their length, resulting from random nicks in the plasmid substrates; this effect is most apparent in the larger fragments at the top of the gel.

dependent synthesis was also observed in the upstream adjacent 52-bp fragment (13% of the total AAF adduct-dependent repair synthesis) and in the downstream adjacent 116-bp fragment (13% of the total). Damage-dependent synthesis in these adjoining fragments is most probably a result of exonucleolytic degradation initiated in both directions from the DNA incisions, followed by resynthesis. There was no increased synthesis in the more distal 57-bp, 127-bp and 139-bp fragments. These data are consistent with an average patch size of 25–40 bp. As expected, addition of 1 µg of 70C antibody to the reaction mixture led to a substantial decrease in the incorporation of labelled nucleotides into the fragment containing the DNA adduct and the two flanking fragments (Fig. 3b, lane 2). Addition of 0.5 µg of SSB with 1 µg of 70C antibody to HeLa-cell extracts resulted in a partial reconstitution of repair synthesis (Fig. 3b, lane 3). This amount of SSB is close to that required to neutralize 1 µg of antibody, so the degree of reversal of repair inhibition is dependent on the absolute amount of SSB in individual cell extracts. When excess SSB (1 µg) was added to the HeLa cell extract in the absence of antibody, an increase in synthesis in the AAF-containing 31-bp fragment and the adjacent 52-bp and 116-bp fragments was evident (Fig. 3b, lane 4). The repair patches in the presence or absence of SSB were indistinguishable in size within the limits of experimental error: with 1 µg of SSB, 73% of adduct-dependent synthesis was confined to the 31-bp fragment, 15% to the 52-bp fragment, and 12% to the 116-bp fragment. Overall, addition of SSB increased adduct-dependent synthesis by about four-fold in each of these fragments. Therefore, the main effect of adding excess SSB to HeLa cell extract was to increase the number of repair events, but not the size of repair patches. Assuming an average patch size of 30 bp, it can be estimated from the amount of radioactive material incorporated (Fig. 2b) that about 2% of repairable ultraviolet-lesions were removed during a standard reaction without added SSB. Addition of 1 µg of SSB increased the number of repaired lesions to about 8%.

How might SSB participate in excision repair? A role for single-stranded DNA binding proteins in the damage recognition step which precedes incision has been suggested by Toulme *et al.*²⁰ on the basis of the relative affinities of bacteriophage T4 gene 32 protein for unmodified native DNA and DNA modified with chemical adducts, including aminofluorene derivatives. Increased affinity for modified DNA has been ascribed to distortions that result in local melting of DNA immediately around the lesion. Local melting also occurs as a result of other kinds of DNA damage²¹, including lesions produced by ultraviolet light. Another possible role for SSB in DNA excision repair is in the recruitment of DNA polymerases onto incised or gapped DNA. This could accelerate the turnover of repair complexes and effectively increase the incision rate, similar to the way that DNA polymerase I and UvrD protein of *E. coli* cooperate to increase the rate of incision by the UvrABC enzyme^{22,23}. If human SSB stimulates DNA repair synthesis by facilitating the binding of DNA polymerase ϵ to substrate DNA, specific protein-protein interactions might be required, which could explain the inability of *E. coli* SSB or adenovirus DNA binding protein to substitute for human SSB.

This work identifies a previously unknown requirement of mammalian nucleotide excision repair *in vitro* for the DNA replication protein SSB. Similar studies with neutralizing antibodies against other known replication factors will be informative in revealing possible additional roles in DNA repair. □

Received 8 November; accepted 7 December 1990.

1. Wobbe, C. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 1834–1838 (1987).
2. Fairman, M. P. & Stillman, B. *EMBO J.* **7**, 1211–1218 (1988).
3. Wold, M. S. & Kelly, T. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2523–2527 (1988).
4. Kenny, M. K., Lee, S.-H. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9757–9761 (1989).
5. Tsurimoto, T. & Stillman, B. *EMBO J.* **8**, 3883–3889 (1989).
6. Kenny, M. K., Schlegel, U., Furneaux, H. & Hurwitz, J. *J. biol. Chem.* **265**, 7693–7700 (1990).
7. Wood, R. D., Robins, P. & Lindahl, T. *Cell* **53**, 97–106 (1988).

8. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, (1980).
9. Wood, R. D. *Biochemistry*, **28**, 8287–8292 (1989).
10. Hansson, J. & Wood, R. D. *Nucleic Acids Res.* **17**, 8073–8091 (1989).
11. Sigghat-Ullah, Husain, I. & Sancar, A. *Nucleic Acids Res.* **17**, 4471–4484 (1989).
12. Hansson, J., Munn, M., Rupp, W. D., Kahn, R. & Wood, R. D. *J. biol. Chem.* **264**, 21788–21792 (1989).
13. Cleaver, J. E. & Kraemer, K. H. in *The Metabolic Basis of Inherited Disease* 6th Edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2949–2971 (McGraw-Hill, New York, 1989).
14. Hansson, J., Grossman, L., Lindahl, T. & Wood, R. D. *Nucleic Acids Res.* **18**, 35–40 (1990).
15. Ikeda, J.-E., Enomoto, T. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 884–888 (1981).
16. Wood, R. D., Lindahl, T. & Robins, P. in *Mechanisms and Consequences of DNA Damage Processing* (eds Friedberg, E. C. & Hanawalt, P.) 57–61 (Liss, New York, 1988).
17. Tanaka, K. *et al.* *Nature* **343**, 73–76 (1990).
18. Weeda, G. *et al.* *Cell* **62**, 777–791 (1990).
19. Nishida, C., Reinhard, P. & Linn, S. *J. biol. Chem.* **263**, 501–510 (1988).
20. Toulme, J. J., Behmoaras, T., Guigues, M. & Hélène, C. *EMBO J.* **2**, 505–510 (1983).
21. Van Houten, B. *Microbiol. Rev.* **54**, 18–51 (1990).
22. Caron, P. R., Kushner, S. R. & Grossman, L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4925–4929 (1985).
23. Husain, I., Van Houten, B., Thomas, D. C., Abdel-Moneim, M. & Sancar, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6774–6778 (1985).

ACKNOWLEDGEMENTS. We thank M. Biggerstaff for preparation and purification of plasmid DNA substrates, and T. Lindahl, M. Meuth and S. Keyse for reading the manuscript.

Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts

David M. Lawson*†, Peter J. Artymiuk*‡, Stephen J. Yewdall*, John M. A. Smith*, J. Craig Livingstone*, Amyra Treffry*, Alessandra Luzzago§, Sonia Levi¶, Paolo Arosio¶, Gianni Cesareni||, Christopher D. Thomas#, William V. Shaw# & Pauline M. Harrison*

* The Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, The University, Sheffield S10 2TN, UK
 § EMBL, Meyerhofstrasse 1, W-6900 Heidelberg, Germany
 ¶ Dipartimento di Scienze e Tecnologie Biomediche, Università di Milano, Ospedale San Raffaele, Via Olgettina 60, I-20132 Milano, Italy
 || Dipartimento di Biologia, II Università degli studi di Roma, Tor Vergata, Via Orazio Raimondo, 00173 Roma, Italy
 # Department of Biochemistry, University of Leicester, Adrian Building, University Road, Leicester LE1 7HR, UK

FERRITIN is important in iron homeostasis. Its twenty-four chains of two types, H and L, assemble as a hollow shell providing an iron-storage cavity^{1–3}. Ferritin molecules in cells containing high levels of iron tend to be rich in L chains, and may have a long-term storage function, whereas H-rich ferritins are more active in iron metabolism^{3–7}. The molecular basis for the greater activity of H-rich ferritins has until now been obscure, largely because the structure of H-chain ferritin has remained unknown owing to the difficulties in obtaining crystals ordered enough for X-ray crystallographic analysis. Here we report the three-dimensional structure of a human ferritin H-chain homopolymer. By genetically engineering a change in the sequence of the intermolecular contact region, we obtained crystals isomorphous with the homologous rat L ferritin^{8,9} and of high enough quality for X-ray diffraction analysis. The X-ray structure of human H ferritin shows a novel metal site embedded within each of its four-helix bundles and we suggest that ferroxidase activity associated with this site accounts for its rapid uptake of iron¹⁰.

Human and rat H and L chains show about 55% identity in amino-acid sequence compared with over 93% identity in the H chains and 83% in the L chains of the two species^{2,4}. Horse spleen ferritin (nearly 90% L), rat liver ferritin (66% L) and recombinant rat L-chain ferritin (rLF, 100% L) give large single crystals with CdSO₄ by vapour diffusion, allowing their three-dimensional structures to be determined^{1,2,8,9,11}. But the addition

† Present address: Department of Chemistry, York University, Heslington, York YO1 5DD, UK.
 ‡ To whom correspondence should be addressed.

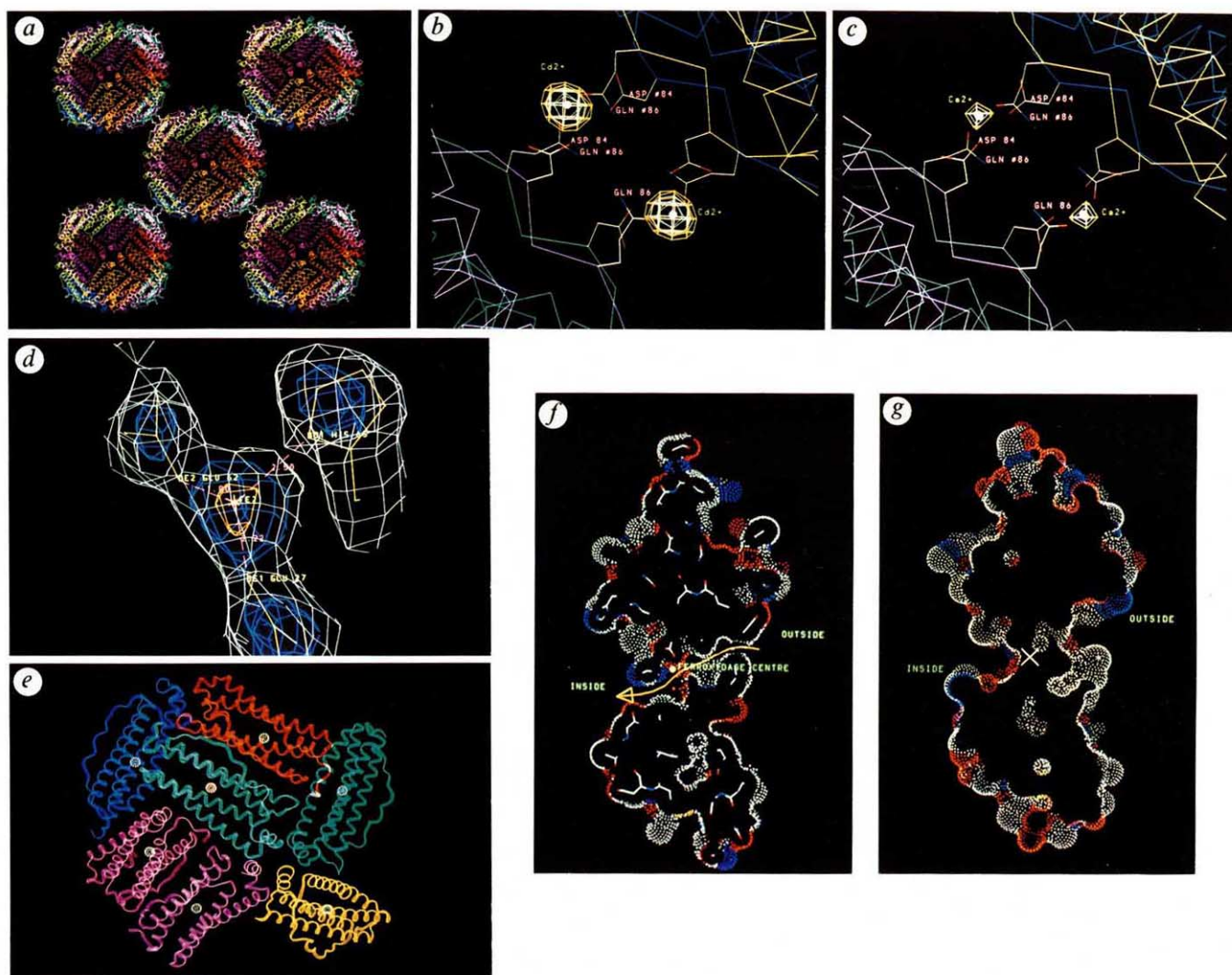


FIG. 1 *a*, Five molecules of rat rLF depicted as α -carbon plots arranged on a face of the cubic unit cell. For clarity a slice of each molecule has been removed such that only 16 of the 24 subunits per molecule are shown. Four of the intermolecular contacts between the central molecule and its neighbours can be seen and these are shown in close-up in *b*. *b*, Contacts in crystals of rat rLF obtained from CdSO_4 solutions. Contacts around 222 symmetry points are through two Cd^{2+} bridges. Cd^{2+} ligands are Asp 84 and Gln 86 situated on long interhelical loops connecting opposite ends of the 4-helix bundles of two different subunits (pink or green and yellow or blue) in each of two molecules. *c*, Intermolecular contacts in crystals obtained from CaCl_2 of a human rHF variant in which an amino-acid substitution Lys 86 $\rightarrow$ Gln has been engineered. The crystals were isomorphous with those of rat rLF and the similarity in the contact region can be seen. Note that Ca^{2+} (18 electrons) is less electron-dense than Cd^{2+} (46 electrons, shown in *b*). Both *b* and *c* show only high electron density ($> 1.5 \text{ e } \text{\AA}^{-3}$) of $(3|F_{\text{obs}}| - 2|F_{\text{calc}}|) \exp i\alpha_{\text{calc}}$ maps, to highlight the metal ion positions. The

Ca^{2+} peaks are two to three times more electron dense than the best-ordered water molecules in the map. *d*, Electron density of the metal and its ligands at the novel metal site in human rHF (the putative 'ferroxidase centre'). The large peak marked Fe^{2+} is probably Ca^{2+} in these crystals. Protein ligands are Glu 27, Glu 62 and His 65 from a single subunit. His 65 occupies slightly different positions in some human rHF variants giving $\text{C}_\alpha\text{-N}_{\text{D1}}$ distances 2.2–3.2 $\text{\AA}$ (our unpublished work). *e*, Seven subunits of the protein shell showing the ferroxidase centre as spheres embedded in the four-helix bundle subunit fold. *f*, Connolly surface diagram of human rHF showing presence of a 1- $\text{\AA}$ channel from the outside surface of the molecule (right-hand side) through the subunit to the cavity surface (left) by way of the ferroxidase centre (solid spheres). *g*, Connolly surface diagram of rat rLF showing an incomplete channel, blocked around the position equivalent to the ferroxidase centre of H chains. The channel of *f* may represent the route by which Fe(II) reaches the centre ($\sim 12\text{--}15 \text{ \AA}$). An alternative route by way of the intersubunit threefold channels is $\sim 60 \text{ \AA}$ long.

of CdSO_4 even at low concentrations to H-rich ferritins (including recombinant human H-chain ferritin, rHF, 100% H) causes instant formation of amorphous precipitates, but not crystals^{6,8}. Examination of the intermolecular crystal contact regions of the horse and rat L-chain ferritins indicates why H-rich ferritins fail to crystallize: in each case the molecules in the crystals are linked by double Cd^{2+} bridges^{1,8,9,11}. These are situated near the crystal two-fold axes, with pairs of aspartate and glutamine side chains from neighbouring molecules forming ligands for the Cd^{2+} ions (see Fig. 1*a* and *b*). In all known H-chain sequences, however, including human, the aspartate (position 84) is conserved, whereas the glutamine (86) is replaced by another amino acid, usually lysine. Although human rHF has not crystallized from CdSO_4 , it has been crystallized from aqueous

polyethylene glycol or methane-pentane diol, but the crystals are seriously disordered, giving poor diffraction and making X-ray crystallographic analysis impossible^{6,8}.

We therefore introduced the Lys 86 $\rightarrow$ Gln substitution into human rHF. This has enabled metal bridge crystal contacts to form, giving crystals diffracting to 1.9 $\text{\AA}$ or better, and allowing the three-dimensional structure to be determined.

The methods used for cloning and overexpressing the rat L ferritin genes and for purifying and characterizing the overexpressed proteins have been described elsewhere^{9,10,12}. The substitution of Gln for Lys 86 was introduced into human rHF and the resulting protein was overexpressed in *Escherichia coli* as in ref. 12. The ferritin variant assembled normally, as shown by gel electrophoresis of the purified protein, and it behaved

TABLE 1 Details of crystallographic analysis

	Resolution (Å)	Reflections measured	Unique reflections measured	Percentage completeness of data	R_{merge}^*	$R_{\text{ref}}^\dagger$
RA	2.3	44,035	10,300	80	0.055	0.195
HA	2.4	54,965	10,378	95	0.128	0.205
HE	2.7	11,444	4,205	75	0.109	—

Because the human H CdM crystals were isomorphous with those of the known structure of L ferritin, solving the structure was very straightforward: the H structure was solved using the phases of L ferritin, and it was therefore unnecessary to search for heavy atom derivatives. Oscillation camera data were collected to 2.4 Å on the SERC Daresbury synchrotron source, the data were processed and then scaled to recombinant rat L ferritin data. The mean isomorphous change (calculated on F^2) was 0.47; this is high, but not unduly so when one considers that 45% of the residues between the L and H structures are different. An initial interpretation of the H ferritin structure was obtained from an $(3|F_{\text{obsHA}}| - 2|F_{\text{obsRA}}|) \exp i\alpha_{\text{calcRA}}$ map and the structure was refined using the restrained least squares program of Konnert and Hendrickson²². Atomic positions and individual temperature factors were refined, the final root-mean-square deviation from ideality for covalent bond lengths being <0.02 Å. All data were used from 20–2.4 Å, with a low resolution form factor modification to allow for disordered solvent displaced by the protein²³. RA, Rat rLF dataset; HA, human rHF dataset; HE, human rHF + Tb³⁺ dataset.

* $R_{\text{merge}} = \sum_h \sum_j |I(h)_j - \langle I(h) \rangle| / \sum_h \sum_j I(h)_j$, summed over all measured reflections.

† R_{ref} , Crystallographic residual, $R = \sum ||F_o| - |F_c|| / \sum |F_o|$.

similarly to the wild type during iron uptake experiments *in vitro*^{8,10,12,13}. It failed to crystallize from CdSO₄, but CaCl₂ yielded large X-ray quality crystals. These proved to be isomorphous (space group F432, $a = b = c = 184.8$ Å) with horse spleen^{1,2,11} and rat rLF crystals ($a = b = c = 184.5$ Å)^{8,9}. Details of the structure analysis are given in the legend to Table 1.

Features of the crystal structure of human rHF and rat rLF are compared in Figs 1 and 2. Figure 1a, b and c shows the metal bridges. Although the metal ion is Ca²⁺ in human rHF (Fig. 1c) and not Cd²⁺ as in rat rLF (Fig. 1b), the octahedral geometry of the metal bridges and the protein ligands are the same, namely Asp 84 and Gln 86, the latter being the residue introduced by mutagenesis in human rHF. The radius of Ca²⁺ (0.97 Å) is very similar to that of Cd²⁺ (0.99 Å).

The human rHF structure has a novel metal site which is not present in rat rLF (or in horse spleen ferritin), although the subunit conformations are essentially the same. Details of this site and its position in the subunit are shown in Figs 1d–g and 2a, b, and the equivalent region of rLF is shown in Figs 1g and

2c. The metal site ligands Glu 27 (of helix A) and Glu 62 and His 65 (of helix B) are conserved in all H-chain ferritins but not L (Table 2) and this might represent the ferroxidase centre¹⁰, previously postulated to lie on the outside of the molecule¹⁴. However, as shown in Figs 1e, f and 2a, b, this site is embedded within the bundle fold of four α helices of each H chain. In L-chain ferritins, residue 27 is tyrosine or histidine (His in rat rLF), residue 62 is lysine and 65, glycine or alanine. The metal site is replaced by a Lys 62 → Glu 107 salt bridge (Fig. 2c). We postulate that a similar salt bridge contributes to the greater stability to denaturants found for human L-chain ferritins (S.L. *et al.*, unpublished).

Figure 1f gives a Connolly surface¹⁵ representation of the subunit (an indication of the molecular surfaces accessible to a spherical probe size set at 1.0 Å) showing the position of the new metal site in human H ferritin. This shows that the metal lies at a distance of 7–8 Å from the inside surface of the molecule and 10–12 Å from the outside. It also indicates the presence of narrow 1.0 Å channels through the 4-helix bundle leading to this site from both the outside and inside surfaces of the molecule. These could provide a means of access from the outside of the molecule for iron and possibly dioxygen. This channel is blocked in rLF (Fig. 1g) by the salt bridge shown in Fig. 2c.

In Fig. 2b we show the putative 'ferroxidase centre' region of a Tb³⁺ derivative of human rHF (Tb³⁺ was introduced because it displaces Fe³⁺ from its initial binding site in horse spleen ferritin¹⁶). In this map, the single metal peak (Fig. 1d) is replaced by an elongated peak which we interpret as two Tb³⁺ ions (marked by crosses) about 3 Å apart. The site marked by the lower cross (site A) is close to the metal site shown in Fig. 2a and has the same ligands with the addition of Glu 107. That marked by the upper cross (site B) is coordinated by Glu 61, Glu 62 and Glu 107. Note that two positions are given for Glu 61. Its density is weak and it seems to alternate between two positions, the second providing coordination for a third Tb³⁺ site (site C). This Tb³⁺ is on the inside surface of the molecule and is also bound by Glu 64. Unlike the other Tb³⁺ sites, a Tb³⁺ ion is found at this same position in horse spleen apoferritin crystallized from TbCl₃ (ref. 2). Indeed, Glu 61, and Glu 64 are conserved in both H and L subunits².

Our current model of ferritin formation (deposition of ferrihydrite inside the apoferritin shell) requires a site for the initial binding and oxidation of Fe(II) which subsequently allows Fe(III) to migrate into the cavity^{17,18}. Site-directed mutagenesis has implicated site-A ligands in the initial oxidation step¹⁰. The relatively sparse ligation at site A would allow subsequent move-

TABLE 2 Partial amino-acid sequences of some ferritins

	A Helix			B Helix			C Helix		D Helix	
	25	30	55	60	65		105	110	140	
	A			B'	B A C A		A + B		A'	
1 HuLi-H	I N L E L Y A	F L H Q S H E E R E H A E K	L H L E K N V	L N E Q V K A						
2 RaLi-H	I N L E L Y A	F L H Q S H E E R E H A E K	L H L E K S V	L N E Q V K S						
3 ChRC-H	I N L E L Y A	F L H Q S H E E R E H A E K	L H L E K N V	L D E Q V K A						
4 TdRC-H	V N M E L Y A	F K E Q S H E E R E H A E K	L Q L E K T V	L E E Q V K S						
5 HuLi-L	V N L Y L Q A	F R E L A E E K R E G Y E R	M A L E K K L	L D E E V K L						
6 RaLi-L	V N L H L R A	F R E L A E E K R E G A E R	L A L E K N L	L D K E V K L						
7 HoSp-L	V N L Y L R A	F R E L A E E K R E G A E R	I V L E K S L	L D E E V K L						
8 RbLi-L	V N L H L R A	F R E L A E E K R E A A E R	L A L E K N L	L D E E V K L						

The three protein ligands of the 'ferroxidase centre' of H chains (Fig. 1d) are indicated by boxes labelled **A** (also included is Glu 107, and together these four residues also form the ligands of site A in Fig. 2b). Ligands of the Tb³⁺ in site B are boxed and labelled **B** and the remaining boxed residue (E64), **C**, is a ligand at site C. Note that some residues ligate more than one site, that residues involved in ligating waters at the Tb³⁺ sites are shown as dashed sites and that residues E61, E64 and E107 are also present in L chains, but E27, Q58 (**B'**), E62, H65 and Q141 (**A'**) are not. 1, Human liver H chain²⁵; 2, rat liver H chain²⁵; 3, chicken red cells²⁶; 4, bullfrog tadpole red cells²⁷; 5, human liver L chain²⁴; 6, rat liver L chain²⁸; 7, horse spleen L chain²⁹; 8, rabbit liver L chain³⁰. Sequences use human H chain numbering. Secondary structure elements are indicated.

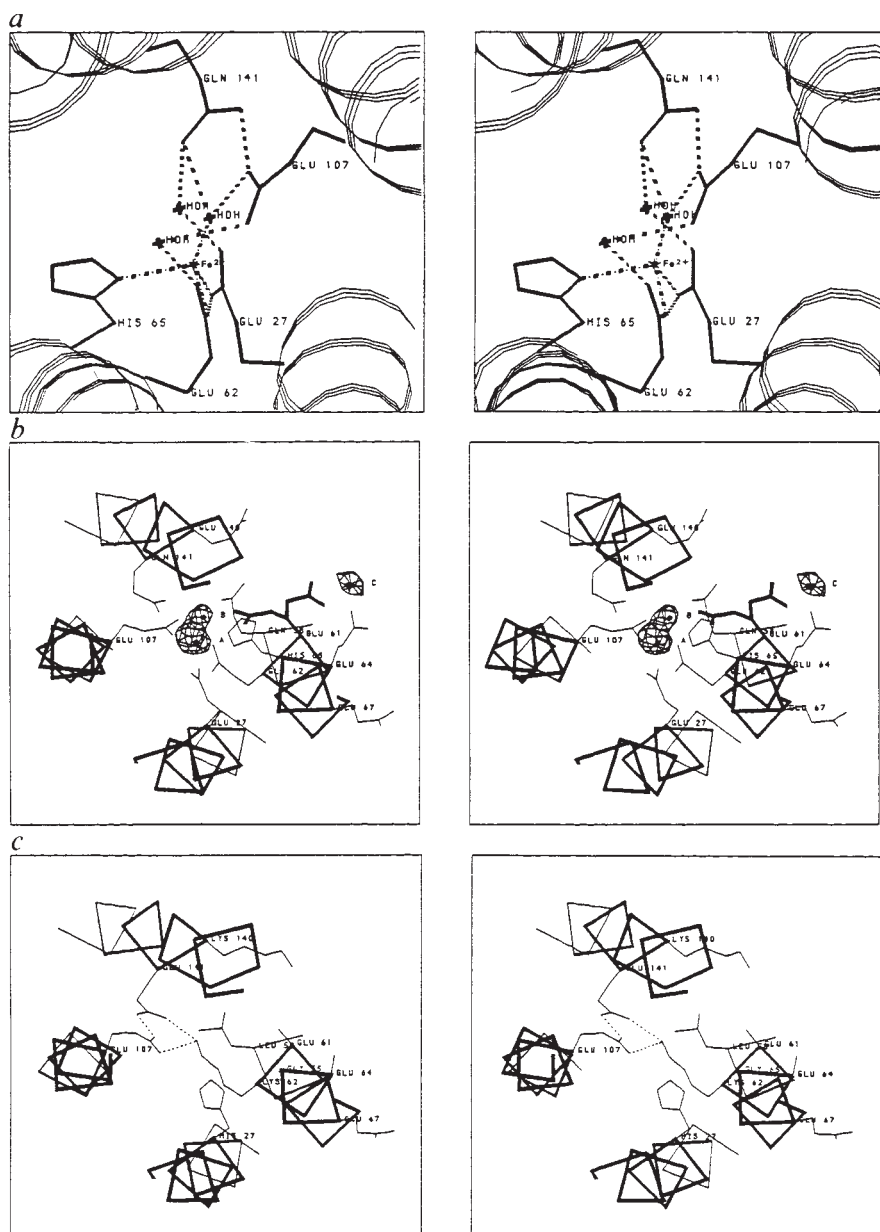


FIG. 2 Stereo diagrams of central regions of ferritin H and L subunits. *a*, Schematic of the metal centre showing its tetrahedral co-ordination geometry including bound water molecules. *b*, Tb^{3+} -binding sites found in human rHF crystals soaked in 10 mM $TbCl_3$. Site A has protein ligands Glu 27, Glu 62, His 65, and Glu 107 and is similar to the metal site of Fig. 1*d*. Site B has ligands Glu 61, Glu 62 and Glu 107 and site C Glu 61 and Glu 64. Site C lies on the cavity surface. Note the two alternative positions of Glu 61. Movement from A $\rightarrow$ B $\rightarrow$ C may represent the pathway of iron from the ferroxidase centre (A) to the cavity (C). *c*, The equivalent region of the rat rLF subunit. The metal ligands Glu 62 and His 65 of human rHF are now Lys 62 and Gly 65. Lys 65 forms an interval salt bridge with Glu 107.

ment of Fe(III). Migration from site A to site C (a possible ferrihydrite nucleation centre)^{1,2,13}, by way of site B, could be facilitated by the movement of Glu 61. But Mössbauer spectroscopic measurements show μ -oxo-bridged Fe(III) dimers as intermediates in iron-core formation¹⁷ and our structural analysis suggests that sites A and B constitute position for such dimers.

The 'ferroxidase centre' in ferritin H chains lies closer to the

middle of the 4-helix bundle along its length than does the binuclear Fe(III) centre of haemerythrin^{19,20}, but is in a location comparable to the iron centre of ribonucleotide reductase²¹ within the four longest helices of the α -helix barrel. The putative Fe(III) dimer lies nearly parallel to the bundle axis as in ribonucleotide reductase, in contrast to the Fe dimer of haemerythrin, which lies perpendicular to this axis. □

Received 31 October; accepted 29 November 1990.

1. Ford, G. C. *et al.* *Phil. Trans. R. Soc. Lond.* **B304**, 551–565 (1984).
2. Harrison, P. M. *et al.* in *Biomimetalisation: Chemical and Biochemical Perspectives* (eds. Mann, S., Webb, J. & Williams, R. J. P.) 257–294 (VCH Verlagsgesellschaft, Weinheim, 1989).
3. Arosio, P., Adelman, T. G. & Drysdale, J. W. *J. biol. Chem.* **253**, 4451–4458 (1978).
4. Theil, E. C. *A. Rev. Biochem.* **56**, 289–315 (1987).
5. Levi, S. *et al.* *J. biol. Chem.* **263**, 18086–18092 (1988).
6. Arosio, P., Gatti, G. & Bolognesi, M. *Biochim. biophys. Acta* **774**, 230–232 (1983).
7. Levi, S. *et al.* *Biochemistry* **28**, 5179–5184 (1989).
8. Lawson, D. M., thesis, Sheffield Univ. (1990).
9. Thomas, C. D. *et al.* *Biochem. Soc. Trans.* **16**, 838–839 (1988).
10. Lawson, D. M. *et al.* *FEBS Lett.* **254**, 207–210 (1989).
11. Banyard, S. H., Stammers, D. K. & Harrison, P. M. *Nature* **271**, 282–284 (1978).
12. Treffry, A., Harrison, P. M., Luzzago, A. & Cesareni, G. *FEBS Lett.* **247**, 268–272 (1989).
13. Levi, S. *et al.* *Biochem. J.* **264**, 381–388 (1989).
14. Bakker, G. R. & Boyer, R. F. *J. biol. Chem.* **261**, 13182–13185 (1986).
15. Connolly, M. L. *J. appl. Cryst.* **16**, 23–28 (1983).
16. Treffry, A. & Harrison, P. M. *J. inorg. Biochem.* **21**, 9–20 (1984).
17. Bauminger, E. R., Harrison, P. M., Nowik, I. & Treffry, A. *Biochemistry* **25**, 5486–5493 (1989).

18. Chasteen, N. D. & Theil, E. C. *J. biol. Chem.* **257**, 7672–7677 (1982).
19. Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. *J. Am. Chem. Soc.* **106**, 618–622 (1984).
20. Sheriff, S., Hendrickson, W. A. & Smith, J. L. *J. molec. Biol.* **197**, 273–296 (1987).
21. Nordlund, P., Sjöberg, B.-M. & Eklund, H. *Nature* **345**, 593–598 (1990).
22. Konner, J. H. & Hendrickson, W. A. *Acta crystallogr.* **A36**, 344–350 (1980).
23. Fraser, R. D. B., MacRae, T. P. & Suzuki, E. *J. appl. Cryst.* **11**, 693–694 (1978).
24. Boyd, D., Vecoli, C., Belcher, D. M., Jain, S. K. & Drysdale, J. W. *J. biol. Chem.* **260**, 11755–11761 (1985).
25. Murray, M. J., White, K. & Munro, H. N. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7438–7442 (1987).
26. Stevens, P. W., Dodgson, J. B. & Engel, J. D. *Molec. cell. Biol.* **7**, 1751–1758 (1987).
27. Dickey, L. F. *et al.* *J. biol. Chem.* **262**, 7901–7907 (1987).
28. Leibold, E. A. & Munro, H. N. *J. biol. Chem.* **262**, 7335–7341 (1987).
29. Heusterspreute, M. & Crichton, R. R. *FEBS Letts* **129**, 322–327 (1981).
30. Daniels-McQueen, S. *et al.* *Nucleic Acids Res.* **16**, 7741 (1988).

ACKNOWLEDGEMENTS. We thank the Commission of European Communities (G.C., P.A. & P.M.H.), the Wellcome Trust (A.T.), the CNR Target Project BTBS (P.A. & G.C.) and the Science and Engineering (P.J.A., W.V.S. & P.M.H.) and Medical Research (P.J.A.) Councils for research grants. P.J.A. is a Royal Society University Research Fellow. Crystal coordinates have been deposited in the Brookhaven PDB, accession number 1HFA.

Two-dimensional electrophoresis

A. Görg

Recent developments in the first- and second-dimension steps of two-dimensional electrophoresis have improved pattern reproducibility.

ALTHOUGH two-dimensional (2-D) electrophoresis based on coupled charge-size fractionation of proteins is a demanding procedure with respect to performance and required expertise, it is, nevertheless, unsurpassed by any other technique for analysing hundreds of polypeptides simultaneously. Therefore, for the comparison of results within laboratories and between laboratories, and for establishing collective protein databases, the importance of achieving maximum reproducibility of spot positions in 2-D patterns cannot be over-emphasized.

In spite of methodological improvements and the availability of sophisticated instrumentation, problems of pattern reproducibility remain when using isoelectric focusing (IEF) with carrier ampholytes. Equilibrium IEF with carrier ampholytes cannot be achieved because of pH gradient instability with prolonged focusing time. As a result of electro-osmotic effects, the pH gradient moves to the cathode ('cathodic drift') and flattens in the centre ('plateau phenomenon'). Consequently time-dependent, not stationary, protein patterns are obtained^{1,2}. In addition, the reproducibility of pH gradient profiles is limited by the batch-to-batch variability of carrier ampholyte preparations. The storage of large amounts of one batch of carrier ampholyte or the blending of new batches in order to maintain a defined pH gradient for long-term studies, although tedious, can help.

With the introduction of immobilized pH gradients (IPGs), the problems of pH gradient stability and reproducibility have been overcome³. Immobilized pH gradients are based on the principle that the pH gradient is co-polymerized, and thus insolubilized within the fibres of the polyacrylamide matrix³⁻⁵. Covalent binding of the IPG eliminates cathodic drift. After focusing proteins at their isoelectric points, constant zone positions over many hours (see Fig. 1), or even over several days, are obtained⁶. For 2-D analysis of basic proteins that are normally lost by cathodic drift, non-equilibrium methods have been devised⁷ in which the sample is loaded at the anode and focusing times are reduced. In this case, basic proteins are retained in the gel, but at the expense of reproducibility. By contrast, when using an IPG in the first dimension, basic proteins are separated under equilibrium

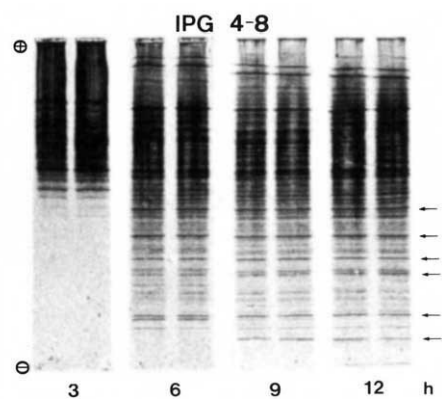


FIG. 1 Time course of isoelectric focusing (IEF) with immobilized pH gradients. Equilibrium IEF is obtained after 9 h. Sample: seed proteins from beans.

conditions and highly reproducible 2-D patterns with distinct spots up to the basic end are obtained (Fig. 2).

IEF in IPG gel strips

The success of using IPGs for 2-D electrophoresis (IPG-Dalt: isoelectric focusing in immobilized pH gradients followed by SDS electrophoresis) depends on their correct use for IEF as the first dimension. After experiencing initial problems in handling IPGs, we have developed a protocol of IPG-Dalt that simplifies existing 2-D procedures, and which is based on a series of experiments with different samples, such as yeast and

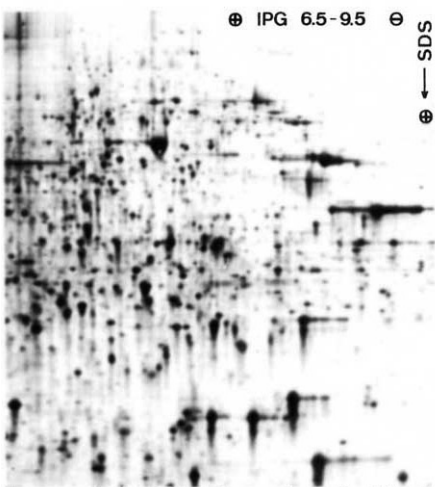


FIG. 2 Horizontal 2-D electrophoresis of basic myeloblast proteins. First dimension: immobilized pH gradient 6.5–9.5 with 9 M urea, 0.5% Nonidet P-40 and 10 mM dithiothreitol. Second dimension: horizontal SDS poregradient gel (4% C, 12–15% T) cast on GelBond PAGfilm.

blood cell proteins, or plant proteins from seeds or tissue⁸. Instead of gel rods, IPG gel strips are used (Fig. 3). This eliminates the tedious and risky procedure of extruding gel rods from glass tubes. Furthermore, IPG gel strips that are cast on plastic backings (GelBond PAGfilm, FMC BioProducts, Rockland, ME) are not only easy to handle but do not stretch or shrink when transferred from the first to the second dimension, which improves consistently the reproducibility of spot position along the IEF axis in the 2-D patterns.

Immobilized pH gradient gel strips (Fig. 3) are prepared by cutting dry IPG slab gels into 3-mm wide strips with a paper cutter. Alternatively, pre-cut IPG gel strips (Immobiline DryStrip Kit, Pharmacia, Uppsala, Sweden) can be used. Before IEF, the dry IPG gel strips are rehydrated to the original gel thickness (0.5 mm) in a mould filled with a solution containing the necessary additives (for example, urea and detergent) for first-dimensional IEF⁹. The rehydrated IPG gel strips are then placed on a flat-bed cooling plate. Samples are applied into silicone rubber frames placed onto the gel

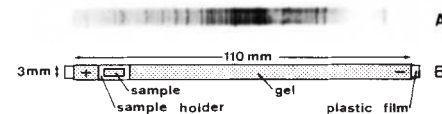


FIG. 3a, Isoelectric focusing in individual immobilized pH gradient (IPG) gel strips. Geometry of the IPG strip. The pH gradient and separation distance may vary by casting different IPGs and/or by using different gel lengths.

surface (Fig. 3b). The recommended sample volume is 20 μ l, although volumes of up to 100 μ l can be applied when the sample cups of the Immobiline DryStrip Kit are used. After IEF, any IPG gel strips that are not used immediately can be stored at -80°C for several weeks before equilibration for the second-dimension run⁸.

Horizontal set ups

For the second dimension, vertical SDS electrophoresis systems can be used. In this case, the equilibrated IPG gel strips are embedded at the top of the vertical slab gels with 0.5% agarose containing electrode buffer^{9,10}. However, IPG gel strips are more easily applied onto horizontal SDS poregradient gels without any embedding procedures⁸.

Horizontal electrophoresis for one- and two-dimensional separations¹¹ became a widely used method with the introduction of ultrathin (≤ 0.5 -mm) gels cast on film supports¹². Reducing the gel thickness from 3 mm to ≤ 0.5 mm, which was made possible by our casting procedure for ultrathin homogeneous¹² and ultrathin poregradient gels¹³ on plastic backings, has eliminated the only remaining disadvantage of horizontal electrophoresis — the creation of a vertical thermal gradient within the horizontal gel layer, which produces distorted protein zones. The introduction of thin, plastic film-supported gradient gels has spurred the development of not only horizontal SDS electrophoresis with ready-made gels, but also horizontal 2-D electrophoresis¹¹.

For 2-D electrophoresis, the SDS gel is placed on the cooling block of a horizontal electrophoresis system. Electrode wicks or buffer strips from polyacrylamide (ExcelGel buffer strip, Pharmacia) or agarose (PhastSystem, Pharmacia) are then applied. The equilibrated IPG gel strip is transferred gel-side-down onto the surface of the horizontal SDS gel along the cathodic wick (strip). Running conditions are dependent upon the gel size used for SDS electrophoresis¹¹.

Horizontal 2-D electrophoresis systems are well suited to the use of ready-made gels. With the introduction of ready-made IPG and SDS poregradient gels, 2-D electrophoresis can be performed without the need for gel casting. Moreover, by using a horizontal electrophoresis unit that provides programmable running and temperature conditions, and staining procedures with precise temperature and time control (PhastSystem), 2-D electrophoresis is becoming an automated operation.

Unfortunately, precast IPG gels to fit the PhastSystem are not commercially available at present. But, even when using laboratory-made IPG gels ($43 \times 50 \times 0.5$ mm³), silver-stained 2-D patterns can be obtained within three hours (90 min for IEF with IPG, 30 min for second-dimensional SDS electrophoresis and about 60 min for silver staining)¹⁴. This set-up has proved to be beneficial for the quick pretesting of unknown samples before applying them to time-consuming, large-scale 2-D electrophoresis.

In conclusion, it can be stated that the reproducibility and handling of 2-D electrophoresis can be improved significantly by using IPG technology in combination with ready-made gels and a horizontal set-up. □

Angelika Görg is Associate Professor of Food Science and Chemistry at the Technical University of Munich, Institute of Food Technology, D-8050 Freising-Weihenstephan, Germany. For more information, fill in reader service number 100.

Migratory trends

A protein visualization system to view protein bands as they separate and a programmable power supply that plots run parameters in real time for troubleshooting purposes — new product highlights for electrophoresis.

PROMEGA is offering the ChromaPhor protein visualization system, a procedure for rapid protein staining during SDS PAGE (Reader Service No. 101). The procedure uses a special gel/buffer formulation in combination with Promega's ChromaPhor green protein stain to stain protein bands while the gel is running. The ability to visualize protein bands as they separate offers several advantages: the quality of the run can be checked continuously; the resolution of specific bands can be optimized as the gel is running; and gel bands can be localized without the need for fixation. As little as 1 μ g of protein per band can be visualized as the gel is running when using 0.75-mm thick minigels, says the company. This level of sensitivity is sufficient for preparative electrophoresis before, for example, sequence analysis, fingerprinting or recovery by elution from the gel. If enhanced sensitivity is required (for example, for analytical applications) post-run enhancement/destaining steps can be performed to provide a detection sensitivity down to <100 ng. Two kits are available with sufficient reagents to cast and stain 20 or 40 mini-gels, costing \$125 (US) and \$195 (US), respectively.

The W.E.P. Company has developed the Post-Transfer System 1000, a semi-automated instrument that is designed to automate the time-consuming, labour-intensive procedures associated with post-



Semi-automated post-transfer system.

transfer techniques such as incubation, staining/probing and overall immunobinding of membranes or filters (Reader Service No. 102). PTS-1000 is a single housing, bench-top unit that consists of a reaction chamber, reagent storage area, reagent delivery system and waste disposal system (a reagent recovery system is optional). The post-transfer system is programmed and controlled through an interface with an IBM/compatible PC and a user-friendly software package. When specifying methods, specific reagents, incubation times and wash cycles can be entered through keyboard commands. Up to 100 sequential methods can be stored. The reaction chamber accommodates multiple gels or membranes of a variety of sizes and the reagent storage area holds six separate reagent pouches, although more can be added. Each pouch holds between 50 and 600 ml of solution. Prices for the basic PTS-1000 unit start at \$4,449 (US). Upgrade options include a built-in incubation processor interfaced with a temperature regulating system and a micro-delivery system.

Thermal Klenow, a dideoxy sequencing enzyme, which is prepared by the partial cleavage of DNA polymerase isolated from a strain of *Bacillus stearothermophilus*, is now available from Stratech Scientific (Reader Service No. 103). The enzyme shows thermal stability up to 75 °C and is intended for use with a variety of sequencing protocols, including the standard Sanger method and the two-step method of Labor and Richardson. Thermal Klenow can be used with either ³²P or ³⁵S. It is suitable for sequencing standard M13 templates, double-stranded templates from phagemid vectors, and is available in 200- and 1,000-unit quantities for £52.50 (UK) and £210 (UK), respectively.

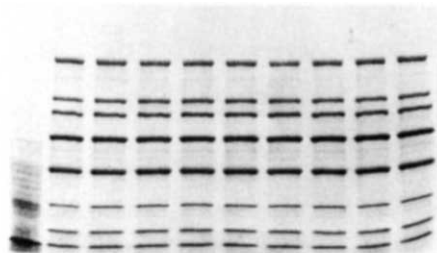
- Chrambach, A., Doerr, P., Finlayson, G.R., Miles, L.E.M., Sherins, R. & Rodbard, D. *Ann. N.Y. Acad. Sci.* **209**, 44–69 (1973).
- Righetti, P.G. & Drysdale, J.W. *Ann. N.Y. Acad. Sci.* **209**, 163–186 (1973).
- Bjellqvist, B. et al. *J. biochem. biophys. Meth.* **6**, 317–339 (1982).
- Görg, A., Fawcett, J.S. & Chrambach, A. in *Advances in Electrophoresis* (eds Chrambach, A., Dunn, M.J. & Radola, B.J.) Vol. 2, 1–43 (VCH Publisher, Weinheim, Germany, 1988).
- Righetti, P.G. *Immobilized pH Gradients: Theory and Methodology* (Elsevier, Amsterdam, The Netherlands, 1990).
- Görg, A., Postel, W., Günther, S. & Weser, J. *Protides Biol. Fluids* **34**, 701–705 (1986).
- O'Farrell, P.Z., Goodman, H.M. & O'Farrell, P.H. *Cell* **12**, 1133–1142 (1977).

- Görg, A., Postel, W. & Günther, S. *Electrophoresis* **9**, 531–546 (1988).
- Görg, A. et al. *Electrophoresis* **9**, 37–46 (1988).
- Strahler, J.R., Hanash, S.M., Somerlot, L., Weser, J., Postel, W. & Görg, A. *Electrophoresis* **8**, 165–173 (1987).
- Görg, A., Postel, W., Günther, S. & Weser, J. in *Electrophoresis '86* (ed. Dunn, M.J.) 435–449 (VCH Publisher, Weinheim, Germany, 1986).
- Görg, A., Postel, W. & Westermeier, R. *Analyt. Biochem.* **89**, 60–70 (1978).
- Görg, A., Postel, W., Westermeier, R., Gianazza, E. & Righetti, P.G. *J. biochem. biophys. Meth.* **3**, 273–284 (1980).
- Görg, A., Postel, W., Günther, S. & Friedrich, C. *Electrophoresis* **9**, 57–59 (1988).

Gel selection guide

The new RNease Gel Kit from International Biotechnologies is an **RNA electrophoresis matrix** that inherently inhibits RNases, resulting in minimal RNA sample degradation (*Reader Service No. 104*). Use of the matrix, which is designed for the linear separation of RNA from 350 bp to 9.5 kb, eliminates the need for the diethyl pyrocarbonate (DEPC) treatment of solutions or equipment. The RNease Gel Kit is designed to replace standard agarose/formaldehyde gels, glyoxal/DMSO agarose gels, methylmercury gels or acrylamide/urea gels. The kit includes Instacryl L, formamide, DTT and a modified Tris borate-EDTA buffer solution in a mix-and-pour format.

Eliminate the need to prepare gel solutions, handle hazardous acrylamide monomer or wait for hand-cast gels to polymerize, with the new Mini-Protean II **pre-cast electrophoresis gels** from Bio-Rad (*Reader Service No. 105*). Designed



Ready-to-use gels from Bio-Rad designed for convenience and reproducibility.

specifically for use with the Mini-Protean II and Mini-Trans-Blot cells, the ready-to-use gels are available in either a Tris-HCl buffer system, for both SDS-PAGE and non-SDS-PAGE gel electrophoresis, or a Tris borate-EDTA buffer system, for nucleic acid electrophoresis and native protein electrophoresis. Each lot of pre-cast gels is quality-control tested for reliability and reproducibility, says Bio-Rad.

According to Schleicher and Schuell, DNA can be isolated from single-cell suspensions, agarose gels or plasmid preparations in less than one hour when using the company's Elu-Quik DNA Purification Kit (*Reader Service No. 106*). The Elu-Quik DNA purification kit eliminates the need to carry out phenol-

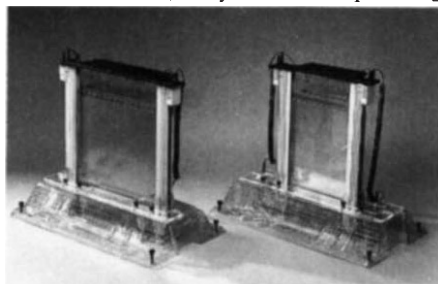


DNA purification in <1 h—see S & S.

chloroform extractions or ethanol precipitations and takes less time than conventional isolation procedures to perform DNA purifications, says S & S. The kit is designed to provide yields of >650 µg of DNA from 10⁸ cells. DNA isolated in this way is suitable for enzyme reactions, cloning or sequencing. The \$165 (US) Elu-Quik Kit contains sufficient reagents for performing up to 200 isolations of <50-kb fragments, 200 isolations of small amounts of genomic DNA (from 10⁶ cells), or eight isolations of large amounts of genomic DNA (from 10⁸ cells).

Electrophoresis apparatus

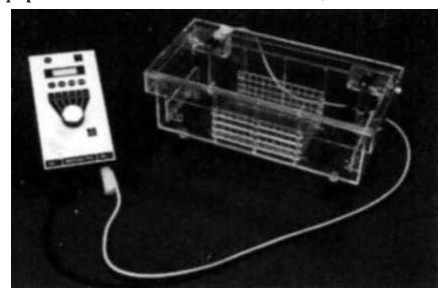
Hoefer Scientific has introduced two new SE 1600 Poker Face II **nucleic acid sequencing rigs**, which are designed to provide smile-free, easy-to-read sequencing



Smile-free sequencing from Hoefer.

gels (*Reader Service No. 107*). An electrically-isolated aluminium plate that is in contact with the gel sandwich provides even heat distribution over the entire gel surface, says Hoefer. The SE 1600 Poker Face II sequencers are supplied in two sizes: the \$850 (US) SE 1600, which can accommodate 33 × 41-cm gels, and the smaller \$725 (US) SE 1650 for 20 × 41-cm gels. These easy-to-assemble units have full-length clamps that slide over the edges of the gel sandwich: no springs or screws are necessary.

Run up to four gels simultaneously, stacked one on top of the other, with the Hybaid Electro-4 **horizontal agarose gel electrophoresis system** for DNA/RNA analysis (*Reader Service No. 108*). To cut sample loading time, the Electro-4 gel system has been designed to coordinate with the 96-well microtitre plate format; gel combs are designed to allow direct loading of samples using 8- or 12-channel pipettes. Users can run mini-, midi- and



Hybaid's four-in-one agarose gel electrophoresis unit.

maxi-format gels in the same gel tank, simultaneously if required. When blotting is performed, pretreatments can be carried out directly in the gel casting tray. The Electro-4 gel system is supplied with four midi casting trays, eight end dams and eight gel combs. Mini and maxi casting trays are optional extras.

OnePhorAll is the new **submarine gel electrophoresis system** from Jordan Scientific that has a built-in buffer recirculation system (*Reader Service No. 109*). By placing the OnePhorAll gel chamber on a magnetic stirrer, buffer is circulated throughout the chamber by an internal pump mechanism. This set-up eliminates the problems of ion depletion, pH changes and temperature gradients, making these gel systems suitable for RNA glyoxal gels, as well as DNA and field inversion gel electrophoresis gels, says Jordan. OnePhorAll is available with medium- and large-sized chambers; both feature no-tape gel casting and multi-length gel beds.

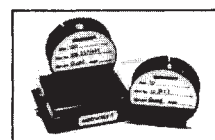
Biometra is offering Agagel for the routine **electrophoresis of small and medium DNA/RNA fragments** using a 10 × 10-cm format and the submerged technique (*Reader Service No. 110*). The £250 (UK) Agagel system has an integral cooling system that is designed to provide the user with accurate temperature control of electrophoresis runs. Other design features include a removable UV-transparent gel tray for viewing gels *in situ*, and a 12-well comb that can be supplied in 1- or 2-mm thicknesses. A preparative 3-well comb is also available.

Power packs

The model 3540 **programmable power supply** from Novex features an output of 3,500 V, 400 mA and 200 W (*Reader Service No. 111*). Run parameters and elapsed or volt hours are shown on the LCD. According to Novex, self-

ADVERTISEMENT

Teflon^{*} Coated and Bare Wire from MEDWIRE



Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

^{*}Reg. TM of E. I. DuPont & Co.

Reader Service No.25

ADVERTISEMENT

Synthetic Oligonucleotides

*Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days*

£4.00 per base
(All inclusive)
0.2 µM scale (approx. 500 µg)

Modified bases, aminoalkylation,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126
FAX XX-44-(0)71 - 733 3877

Reader Service No.56

Custom DNA

*Purified and Shipped in 48 hours,
Worldwide.*

100 ug, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393
In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.26

MERSEYSIDE INNOVATION CENTRE

**TECHNICAL
TRAINING VIDEOS**

Watch the experts at work

**PROTEIN
ELECTROPHORESIS**

by MJ Dunn and JM Graham

Animated sequences to explain the theory
Detailed demonstration of:
Forming SDS Polyacrylamide gels
Electroblotting Quantitation
Gradient gels Isoelectric focusing
2D electrophoresis

MIC Medical Ltd 051 708-0123
131 Mount Pleasant, Liverpool L3 5TF

Reader Service No.10

TIB MOLBIOL BERLIN

CUSTOM DNA SYNTHESIS

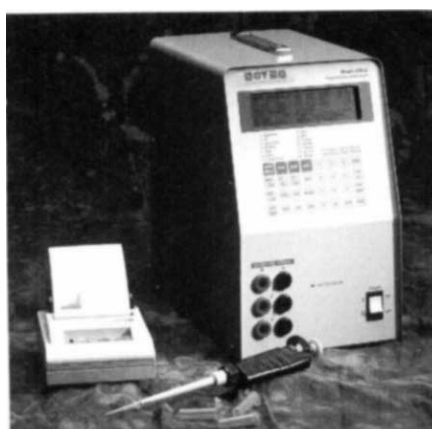
DM per base

5,-

DM 95 - per oligo

FAX (0049) 30 6261960

Reader Service No.79



Novex's power supply with graph screen. prompting menus ease the user through the programming process. The \$2,750 (US) model 3540's memory capacity can store nine run protocols of up to five steps each. A special feature of the power supply is the graph screen, which plots current, watts and volts in real time. This allows data from an entire run to be preserved and recalled at a later date, if there is a need to troubleshoot problems with gels or buffers. A hard copy of the screen display can be obtained with an optional printer. Safety features of the model 3540 are designed to protect against overload, short circuits, open circuits, current leaks and overheating.

For high-voltage vertical electrophoresis of sequencing gels, Stratagene has developed the compact and lightweight FeatherVolt 2000 **volt-sequencing power supply** (Reader Service No. 112). The \$1,195 (US) FeatherVolt 2000 power supply can be operated in three modes: constant voltage, constant current, or constant power, with automa-



tic crossover among all three. A four-digit LED monitors voltage, current and power. The power supply has a maximum output of 2,000 V, 100 mA and 200 W, says Stratagene. Each mode is continuously variable and the output is voltage regulated. As a safety measure, the power supply is short-circuit protected.

Picture perfect

Polaroid's DS-34 camera, a **portable self-contained imaging device**, provides users with an instant record of protein and nucleic acid gels (Reader Service No. 113).

The DS-34 uses 3.25 × 4.25-inch black and white or colour film, or full colour overhead transparencies for record keeping and presentation purposes. A positive/negative instant film is also available, which can be used for producing publication-quality output. No focusing is required with the DS-34. The main feature of the camera is the range of interchangeable hoods (from 100 to 200 mm) that are available, which are compatible with most ultra-violet and white-light transilluminators. The £250 (UK) DS-34 imaging system is supplied with a hood, a filter to ensure jet-black backgrounds, and a starter pack of high-speed instant black and white Type 667 film.

Joyce-Loebl has improved the TV-input capability of Gemini, the company's integrated system for the **analysis of one- and two-dimensional electrophoresis gels, blots and TLC plates** (Reader Service No.



Gemini — an integrated 1- and 2-D analysis system for gels, blots and TLC plates.

114). These latest modifications to Gemini provide a spatial resolution of 1,024 × 1,024 pixels, coupled with 256 grey-level resolution for improved spot discrimination in images captured through the TV camera, says Joyce-Loebl. To maximise the accuracy of the captured images, facilities for shading correction and image averaging are provided. The upgraded Gemini system uses a faster algorithm for spot detection and has an improved 2-D analysis capacity, which has doubled the number of spots that can be measured per gel and increased the number of gels that can be compared per experiment. Other design modifications include: automatic file sequence naming, modified menu layouts and improved data output presentations. The introduction of MIDAS (Multi-line 1-D Analysis Software) provides users with the capability of matching and quantitatively comparing multiple profile data sets. Prices for the Gemini system start at about £15,000 (UK). □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.